# Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 21 (2019), Issue 7 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

## Contents

### Viewpoints



### Reviews

## Original Papers

## Corrigenda and Addendas

Viewpoint

# Data Work: Meaning-Making in the Era of Data-Rich Medicine

Amelia Fiske[1,2], BA, PhD; Barbara Prainsack[3,4], MA, Dr Phil; Alena Buyx[1], MD, PhD

[1]Institute for History and Ethics of Medicine, Technical University of Munich School of Medicine, Technical University of Munich, Munich, Germany
[2]Department of Anthropology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[3]Department of Political Science, University of Vienna, Vienna, Austria
[4]Department of Global Health & Social Medicine, King's College London, London, United Kingdom

**Corresponding Author:**
Amelia Fiske, BA, PhD
Institute for History and Ethics of Medicine
Technical University of Munich School of Medicine
Technical University of Munich
Ismaninger Straße 22
Munich, 81675
Germany
Phone: 49 8941404041
Email: a.fiske@tum.de

## Abstract

In the era of data-rich medicine, an increasing number of domains of people's lives are *datafied* and rendered usable for health care purposes. Yet, deriving insights for clinical practice and individual life choices and deciding what data or information should be used for this purpose pose difficult challenges that require tremendous time, resources, and skill. Thus, big data not only promises new clinical insights but also generates new—and heretofore largely unarticulated—forms of work for patients, families, and health care providers alike. Building on science studies, medical informatics, Anselm Strauss and colleagues' concept of patient work, and subsequent elaborations of articulation work, in this article, we analyze the forms of work engendered by the need to make data and information *actionable* for the treatment decisions and lives of individual patients. We outline three areas of data work, which we characterize as the work of supporting digital data practices, the work of interpretation and contextualization, and the work of inclusion and interaction. This is a first step toward naming and making visible these forms of work in order that they can be adequately seen, rewarded, and assessed in the future. We argue that making data work visible is also necessary to ensure that the insights of big and diverse datasets can be applied in meaningful and equitable ways for better health care.

***(J Med Internet Res 2019;21(7):e11672)*** doi:10.2196/11672

**KEYWORDS**
big data; data work; medical informatics; internet; data interpretation; decision support systems

## Introducing Data Work

With health care becoming increasingly data driven, more and more domains of people's lives are *datafied*, that is, they are translated into a format that lends itself to automatic processing and computation. Examples range from data generated by individuals using health and lifestyle smartphone apps, the digitalization of health records, data from direct-to-consumer testing or drug trials, to biobanking research and clinical genetic testing. Data from increasingly diverse sources are thus rendered, at least in principle, usable for health care purposes. Yet, deriving insights for clinical practice and individual life choices, and deciding what data or information should be used for these purposes, poses difficult challenges. Indeed, it has been argued that "big data won't cure us" [1]; turning data into meaningful information for clinical practice requires tremendous time, resources, and skill. Thus, big data not only promises new clinical insights but also generates new—and largely unarticulated—forms of work for patients, families, and health care providers alike.

Building on insights from science studies, medical informatics, as well as on the concept of patient work and subsequent elaborations of articulation work [2-4], in this article, we analyze the forms of work engendered by the need to make data and information *actionable* in the health care context [5]. Doing so brings the perspective of social and ethical studies of biomedicine into conversations around digital medicine, emerging technologies, medical devices, apps, engineering, and informatics. We outline 3 areas of data work, which we characterize as the work of (1) supporting digital data practices;

(2) interpretation and contextualization; and (3) inclusion and interaction. We argue that it is necessary to name and make visible these forms of data work for them to be adequately acknowledged, assessed, and rewarded. Making data work visible can also help to ensure that the insights of big and diverse datasets can be applied in meaningful and equitable ways for better health care. Although this paper primarily aims to highlight emerging forms of work in the era of data-rich medicine that have not been explicitly or comprehensively considered heretofore, we close by outlining avenues for future practice and policy.

## *Data Work: A Persistent Challenge for the Era of Data-Rich Medicine*

### Emerging Forms of Data Work

Controversies surrounding data use, storage, and sharing illustrate the important ethical questions that emerge when data collection and analyses are applied to new ends. Examples in the news abound, for instance, the rise of direct-to-consumer genetic testing for diseases such as cancer, seen recently through the example given on National Public Radio of an uninsured American woman concerned about her risk of breast cancer [6]. Upon reading her results from 23andMe, the woman admitted feeling less urgency about getting additional testing or mammograms with her physician—something that geneticists worry could pose problems for individuals carrying variants undetected by tests offered by commercial sources, or for those who receive summary advice from individuals without proper training, possibly leading to clinical harm in the future. Other disputes have emerged when health technologies are applied to new ends, such as the recent identification of the *Golden State Killer* in California, United States, in April 2018. Detectives were able to identify the perpetrator by matching crime scene evidence with a family member's DNA profile that the family member had uploaded to a genealogy website. The incident, and subsequent admission that private companies have shared access to their database with law enforcement to find potential suspects, spurred controversy among experts and the public over the legitimacy of using the personal data of volunteers who had not consented to such law enforcement applications [7,8]. Controversies such as these—as well as others surrounding privacy and, for example, the hacking of medical devices [9], or matters of justice and fairness in algorithms [10]—point to the centrality of data at the heart of negotiations over the public good; the status of data generated outside of official forums of science and medicine; and central ethical questions of privacy, consent, and benefit that are emerging in new configurations [11,12].

By *data work*, we are referring broadly to the forms of technological, analytical, and emotional work undertaken by all actors within the health care system that is necessary to make data clinically and personally meaningful. Here, we focus on the emerging forms of data work undertaken by patients and health professionals. This work is already occurring, for example in the interpretation of direct-to-consumer genetic tests [13], efforts to improve patient understanding of broad consent in biobanking [14], or as researchers define proteomic markers of risk, such as for ovarian cancer [15], albeit in an often unrecognized and patchwork manner. Although science studies scholarship has explored various determinants and conditions of data production in the health sphere [16-18], the types of work that are necessary to make diverse forms of health data actionable in daily life by patients and health professionals have not been systematically addressed or conceptually analyzed [19]. *Data work* is ongoing and constitutes a formidable yet underresearched challenge in the era of data-rich medicine. But what kinds of work does this entail, and for whom? What divisions of work or tools would be necessary for addressing ethical and equitable applications of data in everyday life?

Empirical studies examining the organization and structure of medical work from a sociological perspective [20-24] have been helpful to draw attention to the often invisible contributions that patients and their family members make to all aspects of health care. However, conceptualizations of such *patient work* in the era of data-driven medicine are, as of yet, largely missing [25]. As debate grows in medicine over how to best actualize voluminous and diverse data for better outcomes in health care [26,27], many of the biggest challenges are of a social, rather than technical, nature [1]. In this context, more systematic attention to the ways in which professional and nonprofessional actors within the health care system help, for example, to create and interpret data, would fill an important gap. In the following section, we outline and describe three areas of emerging forms of work that have accompanied the turn toward big data in medicine, identify who does this work, and sketch potential ways of addressing concerns that arise in connection with this work. For each area of data work, we offer one vignette to illustrate the forms of data work that are already ongoing. Although the boundaries between these different types of data work are fluid, we posit that there is analytic value in drawing out the key features that characterize each activity to see what challenges they pose and how we might address these.

### More Than a Click Away: Supporting Digital Data Practices

> *G is excited about a new app that promises to keep track of his heartbeat, steps taken, and minutes slept, and to aggregate these data with his weight, blood pressure, and glucose levels. Yet, after looking at the Terms of Service, he realizes that by using the app he signs the rights to his data over to the company. G wonders if there is another option. Finding himself mired in pages of legalese, he starts to think, "maybe I'm just too uptight—what could they really do with all this data, anyways?"*

Advances in mobile devices have changed how health information and support services are being accessed, communicated, monitored, and acted upon [28], offering potential gains ranging from clinical oncology [29] to improving health outcomes for low-income populations [30]. As a result, patients create and engage with health data not only in medical institutions but also in their homes and in other places outside the clinic, via wearable or portable devices, or other tools. Patients and health care professionals alike are faced with ever wider types and larger volumes of data that could *potentially* be relevant for health care, without a clear understanding of the implications of specific forms of personal data [31]. In the domain of mobile apps, one form of emergent data work is the work done by patients who search through the fine print of Terms of Services of new devices and apps to decide whether or not to use them. This is often not easy to do; for instance, the interests of a company providing a digital health device or platform might be hard to fathom for a user, posing potential concerns for individuals who are consenting to data use agreements for a health app, or uploading their medical history to a Web portal for a rare-disease patient community.

Furthermore, the ability to learn about genetic traits—which can now be done with ever lower expense on the internet—raises profound ethical challenges. As Kung and Wu ask, “if we discover certain genetic risk factors in our genome sequences, do we (or our health care providers) have a responsibility to inform our family members who might have similar genetic risks?” [32]. Privacy matters, and the effects of new health technologies on future generations all become important concerns with which individuals have to grapple, while very little or no guidance may be available. The work that people are doing when navigating the landscape of available offers, and in deciding what test they should take and what behavior they should track in an attempt to maintain or increase their health, should not be trivialized. There are increasing expectations that individuals make informed decisions as responsible managers of their health, and now also as owners (morally or legally) of their data.

In addition to such new data work for patients, another novel form of data work emerges for health professionals. This consists of assisting patients and their families in navigating the landscapes of available offers for tests, devices, and services, and helping them to decide whether they should *datafy* certain aspects of their lives and bodies in the first place. This data work includes engaging patients in conversations about the implications of their potential data contributions before patients have had practical experience with these digital practices, and about whether and how they should consider engaging in certain activities. Steering patients through the multitude of options is an important yet complex task. Recent studies have also shown that socioeconomic status, age, English literacy, and digital literacy all play important roles in the uptake of new mobile technologies such as health apps [28,29,33] in engaging in Web-based participatory medical research [34] and in efforts to counter the *digital divide* [35-37]. Importantly, these differences also influence whose data are missing from the broader evidence base upon which future decisions in medicine might be made [25]. This points to the growing need to ensure that such digital health practices and technologies do not exacerbate existing inequalities in society or health and the critical role that health professionals are called upon to play in mediating digital engagements.

Looking forward, we thus anticipate that the data work of professionals in this space will include not only assisting patients in navigating this digitalized network of health-relevant services but also assisting those who cannot, or choose not to, engage digitally [38]. As noted, people who do not make use of digital tools to collect, view, and share data and information about themselves can become *missing bodies* in today’s health care environments, meaning that their bodies, needs, and behaviors remain unaccounted for in decisions made on the basis of new digital health sources [25,39]. Especially when the stakes are so high, neither offering guidance on patient use of digital tools and new health products nor understanding the advantages and disadvantages of the many new products on the market every day is intuitive. To be effective, these activities require time and appropriate training, which are in very short supply in today’s time-starved health care environment [40-43].

One possibility, as we have argued elsewhere, to better support both patients and providers in the era of data-rich medicine would be the creation of a new, intermediary profession entirely, which we have termed *health information counselors* (HICs) [44]. With a broad knowledge of various kinds of health data and data quality evaluation techniques, as well as analytic skills in statistics and data interpretation, our vision is that HICs would be trained also in interpersonal communication, health management, insurance systems, and medico-legal aspects of data privacy. Operating as a clinical consultancy, HICs would have the ability to translate the complex language of data into intelligible and actionable information for both patients and physicians. The creation and implementation of such a specialty would enable patients to make educated, truly autonomous choices about how these novel forms of health data can inform their personal care decisions. Although certainly not the only option for addressing the aforementioned concerns, the creation of this new specialty would go a long way in assisting individuals such as G from our opening vignette, as well as health care professionals, to consider their options and make more informed choices about how increasing amounts of health data and information can or should inform health care.

## How to Tell It All Apart: The Work of Interpretation and Contextualization

> *A brother informs his sister, L, that he has done a commercial DNA test that revealed that he could be a carrier for a particular condition. Because L is considering having a child with her partner, she wonders if she should undergo testing, and what this would mean for their decisions going forward. In reading the leaflet provided by a company offering the testing, she is not sure what is meant by the information that carrier reports may vary in detection accuracy by ethnicity (L has Ashkenazi heritage), and that carrier testing does not include all possible variants for a given condition. L wonders: “What would this information mean for me personally? Who could I ask about this?” She is unsure if her primary care physician is the right person to ask, and who else she could turn to.*

Testing practices such as the one described in this vignette have become a means through which individuals understand themselves and their relationship to society. For some patients, the quantified self can allow people to see new patterns or make changes in their lives: counting steps might lead one to take the stairs, and tracking sleep patterns might lead another to try and get an extra hour of sleep. For others, finding out the percentages of one's global ancestry or likelihood that they could be a carrier for a genetic condition represents personally significant information. Yet, the effects of health-related data and information are often difficult to anticipate and understand. Randomized controlled trials have studied the clinical impact of patients' use of mobile and digital health tools, such as the effectiveness of smartphone apps for weight loss and self-applied therapies [45-47]. Other studies have shown the necessity of looking at patient experience of digital tools to understand how mobile health affects self-management of chronic conditions or changes in well-being [48-50]. In some cases, certain forms of health information may have personal utility for some people even if they lack clinical utility [51]. Overall, such research shows that further work—such as prescreening or offering hands-on assistance and consultation—is needed to turn a health app or Web-based service such as direct-to-consumer testing into a meaningful tool for an individual patient [52].

Data science holds the potential to offer important predictive and diagnostic information that can be used to improve decisions taken by clinicians to reduce error or support estimates, such as the likelihood of medication adherence or organ rejection [53,54]. Yet, from body temperature to steps taken, heartbeats, and hydration levels, it is not yet clear what the biometric data collected via devices such as wearables or smartphones will mean for medical practice and health practitioners. The same is true for nonmedical grade testing services. Both the quality of the data and the possibilities of data interpretation are relevant here. Commercial devices are often not calibrated to the standards of medical grade devices, particularly if not used exactly as intended, which means that data collected through them cannot be used as reliable evidence for health care decisions. Internet communities and apps that offer peer-to-peer support can also be problematic when inaccurate or purely anecdotal information is shared, for example, how-to-hack Web-based tutorials or the increasing use of YouTube as a platform for disseminating misleading health information or offering problematic interpretations of existing data on conditions such as anorexia and bulimia [55-58].

The complex task of discerning irrelevant, unreliable, or misleading health information from relevant, valid, and clinically actionable personalized health resources and then interpreting and contextualizing these for specific patients and their families is emerging as a significant, and time-consuming, activity for health care providers. In our survey of health professionals working in the region of Schleswig-Holstein, Germany, providers expressed repeated concerns about the increasing amount of time devoted in patient encounters to explaining why data from a Web-based genetic test are not relevant, or why a novel therapy reported on a patient community website is not the best choice for a family member [59]. These findings are echoed by recent reports that have pointed to the need for new and improved decision aids to situate the most personally relevant and high-quality digital tools for patients [28,60]. Although some standardization work regarding this issue is currently undertaken by groups such as the Consumer Technology Association, the creation of new devices, apps, and programs and the demands these pose regarding data interpretations and contextualization continues to exceed regulatory processes and physician workloads.

In this context, data work includes deciding which data or information are reliable and relevant for a given context of a specific patient—including contexts outside of the clinic—to decide which intervention, tool, or device might be appropriate or helpful in a given situation, or in future. Again, this is a complex task. For example, discerning whether data brought in by patients derived from commercial or hacked devices can be clinically relevant involves researching devices, analyzing the information they collect, and deciding if, and how, the information generated could be used to inform individual case decisions. In some instances, such data work could include contacting the company producing the device for more information, or seeking out additional resources to evaluate the reliability of the data generated. The same is true for commercially available genetic testing, or the results derived from nonstandard forms of research occurring on patient platforms, such as in some citizen science initiatives [61].

The work of contextualization also increasingly extends to the analysis of the algorithms used to produce data in the health care context. Algorithms are neither 'objective' nor intrinsically neutral and they can exacerbate societal inequities. Biases—regarding race, gender, educational status, body mass index, and so on—are programmed into systems, and the characteristics of datasets that these systems use to learn might reproduce inequities [10,62]. As more and more parts of our lives are being *datafied*, there is an increasing need for contextualization of the health data gained through Web-based tests, mobile, and digital technologies [63]. This includes making the context of data explicit, and asking questions such as: What data was collected, from whom, and how? What do these data represent, and what do these leave out? How has it been made legible for computation, and what has been lost or gained in the process? Such questions are increasingly necessary given the growing ubiquity of domains of everyday life being understood through computational practices. All of the above forms of evaluation require a significant degree of analytical and computational literacy and reflection on whether a particular process of meaning-making relies on evidence that is accurate and reliable in a technical sense, if it is mostly personal and social, or if it is indeed faulty or misleading [64].

Patients, in addition to health care professionals, are also increasingly participating in specific forms of work, including outside of clinical settings. This is the case, for example, when patients do internet searches and seek assistance in making sense of reports or articles found on the internet, thus engaging in the work of sorting, interpreting, and analyzing diverse and often competing sources of information. Often this type of work is undertaken by family members or caregivers to support a patient's health care choices. The work of contextualization will remain a persistent challenge in years to come as more devices, apps, health-related services are offered to individuals outside the supervision of medical professionals. As an area that is in need of robust investigation and public debate, it would be productive to have greater involvement by scientific and academic societies in conducting and sharing analysis of how data can and should be used. Although some of this work is already ongoing, such as recent reports addressing the opportunities, risks, and ethical questions associated with use of *good* artificial intelligence (AI) in health care, or developing specific suggestions that can be taken up by stakeholders and policy makers at national and international levels [65,66], further work is needed on different aspects of the use of big data in medicine. By fostering greater debate, and providing material that is available for lay readership to engage with the stakes of their data engagement, academic scholarship can better support digital literacy in this area.

### Facilitating Conversations About Aims and Interests: The Work of Inclusion and Interaction

> *Upon entering the hospital for an inpatient stay, P, an elderly patient, is asked to opt-in to the institutions' efforts to improve efficiency and calculate predictive health and frailty scores for patients [67]. P is not sure what this means, or how his personal information will be stored and used in the future. [67]*

The prior areas of data work that we have outlined have emphasized the need for a strong awareness of what new data, tests, and technologies are available and how they work. Data-rich medicine highlights a number of ethical issues [11], not least of which is the cross-cutting work of addressing different aims, goals, and interests. As data are increasingly accessible, distributed, revealing, and reidentifiable, ethical concerns pertaining to digital health, large datasets, and precision medicine are multiplying, including issues of consent, protecting participant privacy concerns, and maintaining public trust [68]. Given that many of data-driven practices track new territory in health, questions of power asymmetries and social-economic value are emerging with new relevance [12,69]. An important form of data work thus involves fostering conversations with and across stakeholder groups around these concerns.

As precision medicine moves away from *one size fits all* approaches to treatment, machine learning approaches are increasingly improving the ability to target patients for specific treatments, such as in the use of DNA methylation to subclassify tumors of the central nervous system [70]. The potential of this work to improve personalized therapies through the use of mathematical models is great, yet both the perceived benefits and the social, economic, and health-related concerns vary by actor [71]. In other words, a provider will likely have a different set of investments in the technology, research, and treatment outcomes than a given patient, a hospital chief executive officer, a pharmaceutical company, or an interested member of the public. A patient might be most concerned about loss of privacy, discrimination, or stigmatization (albeit also interested in disease prevention and better treatment), whereas company representatives might be uneasy about losing exclusive access to datasets and find themselves at odds with community members committed to principles of open access. Thus, a central aspect of data work is creating the spaces for interaction and facilitating conversations between differently motivated parties, such as assisting one actor to understand the concerns of another, or finding novel ways to address specific concerns around discrimination, privacy, or equity.

In the digital era, privacy concerns take on a different configuration than in the paper age [72]. Data work in the context of privacy is not limited to simply informing patients of what happens with their data and information once it has been collected but includes moving beyond the widely accepted ethical principle of respecting patient autonomy [73] to including patients in decisions over what type of information will be collected about them in the first place, and to what end. The General Data Protection Regulation (GDPR) introduces protections that began in 2018 across the European Union (EU; including the United Kingdom), but outside of the EU, there is little agreement on regulatory standards for digital health tools or data protection in research, databanks, and big data [61,74-77]. Despite the overall objective of European harmonization, the GDPR gives member states leeway, for instance, in determining whether patient consent is required for secondary data use in medical research, and in which form [74,78]. These national differences have various practical and normative consequences, most of which have not yet been fully analyzed, as well as different implications for research practice across member states. Legislation in countries where data protection is sector specific, rather than general, such as Health Insurance Portability and Accountability Act (HIPAA) in the United States, has addressed data privacy and security concerns relating to medical information since 1996. Subsequently, the HIPAA omnibus rule of 2013 modified the Act to meet guidelines set by the Health Information Technology for Economic and Clinical Health in 2009. Such efforts have expanded the extent of HIPAA beyond providers and insurance companies to also consider the role of business associates. However, even though concerns surrounding patient privacy and the reuse of health information have long been an important topic, the ability of existing regulation such as the GDPR or HIPAA to fully address the concerns emerging in the age of big data remains unknown [79]. We highlight here that the forms of data work we identify can pose particular challenges for privacy, including: the rapid rate of digital innovation; that decisions need to be made on both on the individual and societal level about which aspects of everyday life should be captured by data in the first place; that harm can occur from data use that is not necessarily illegal [80]; as well as broader concerns about data privacy protection legislation.

How to effectively engage a range of stakeholders, including patients, providers, researchers, and insurance companies in these data work concerns, is an ongoing discussion in both clinical practice and biomedical research [81-83]. One critical area of data work for health care providers and researchers is holding conversations with patients about data collection and privacy to better understand the impact of collecting anonymized patient health data in research [14,84]. Data work includes ensuring that patients are party to the decisions about what information will be included in their records, who the gatekeepers for this information are, and for which goals and for whose benefit this information will be used beyond the realm of individual-level health care decisions. It is critical that these discussions include reflections on how data could potentially be reused in the future, for example, the use of predictive health and frailty scores by insurance companies as mentioned in the vignette, as well as the identification of potential protections to guard against uses of data that could be harmful or exclusionary to patients. Specific conditions of access, reuse, and reidentification need to be identified and continually updated in light of new digital advances.

In particular, digital technologies raise important questions over the access of personal information. Each patient's needs and interests are influenced by their human, natural, and artifactual environments. An individual's decision to access his or her electronic health records or use a Web-based genetic testing service is not just a choice made by an atomistic individual but an act shaped by the person's family ties and social relations, his or her connection to others, and the country in which he or she lives [85]. For example, an individual may want to share and discuss this health information with his or her partner or children [82]. This decision to share and discuss information received is not an afterthought but may well have shaped the decision to obtain information in the first place [86]. This layer of dyadic or multilateral forms of decision making can vary significantly across cultural contexts.

In sum, joining distinct datasets from different types, locations, and ethical standards adds additional layers of deliberation to well-rehearsed ethical considerations. Recognition and fostering dialog around aims and goals and the more complex, potentially shared nature of decision making in the era of big data is a critical form of data work. However, how this can be achieved when data are held in dispersed locations and are diverse in nature is entirely unclear. It will require close communication between the patient and the health care provider to ensure that the built-in decisional pathways offered by data-driven practices do not eclipse individual priorities. One potential way of addressing this concern is to reconsider existing methods for ensuring patient privacy and protection and addressing them through regulatory measures, for example through the GDPR in Europe. According to the GDPR, for personal data to be processed lawfully, either individual consent is required, or a legal authorization has to apply. The most relevant legal authorization in the medical context is the research exemption (Article 89). However, particularly in view of international research collaborations, further work is necessary on how GDPR is implemented across individual countries. To provide an example, in line with Article 89, Germany now allows data processing of pseudonymized data for scientific or historical research purposes or for statistical purposes, at least prima facie, without requiring individual consent. However, neither clear guidance exists as of yet for how these purposes are exactly delineated nor have studies been conducted on how this new legal provision has penetrated research practice and how effects differ from countries that are more restrictive. Countries that have long-term experience with more permissive approaches, such as broad or blanket consent (eg, the United Kingdom) and the processing of genetic data should help to anticipate the implications of the novel practice and to raise the standards for how informed consent can be better operationalized in light of the concerns of big data—also in areas outside of Europe [87].

## *Who Does Data Work: Patient Work 2.0?*

The different kinds of technological, intellectual, social, and emotional work sketched here mean that patients, their families, caregivers, and other health care providers will be faced with an increasing range of tasks in the domain of health care, which we have summarized in a list (Table 1). This list of tasks is not meant to be exhaustive but rather to make explicit some of the principal kinds of work involved in making data matter medically. Many of these concerns overlap; we expect that new forms of expertise will continue to emerge along with clinical and technological advances.

**Table 1.** Outline of various types of data work with examples.

| Types of data work | Why is this work needed? | Examples of data work in practice; ongoing and possible in the future |
|---|---|---|
| Supporting digital data practices | Engagement with health data is increasingly taking place outside the clinic, and it can also create digital divides; traditional means of managing and evaluating data are increasingly not suited to meet the realities of the digital age; persistent difficulties in assessing accuracy and appropriateness of diverse, unvalidated forms of health data. | Patients research and consider the implications of data; health practitioners assist in navigation of data relationships; creation of guidelines for how to evaluate new digital technologies or assess internet sources; identification of how digital interaction can create new patterns of exclusion. |
| The work of interpretation and contextualization | Unclear what biometric data collected via devices such as wearables or smartphones will mean for medical practice; misleading or false health information is often shared on the internet; the algorithms that produce data are neither objective nor intrinsically fair; the full implications of diverse, unregulated health information are often difficult for users to discern or anticipate. | Expert guidance on how to decide which devices and resulting data are reliable and relevant for a given context; research on reliability of commercial devices; provision of prescreening and assistance to make digital health tools meaningful for individual patients; identification of biases built into algorithms of datasets, devices, and models. |
| The work of inclusion and interaction | Data are increasingly accessible, distributed, revealing, and reidentifiable, creating new ethical concerns; perceived benefits of the data-driven medicine and the social, economic, and health-related concerns vary by actor; patient experience of digital tools affects self-management of chronic conditions and well-being. | Support for patients in determining their priorities, needs, and wishes with regard to their digital health activities and data collection and use; facilitation of conversations between differently motivated parties about aims, goals, and interests. |

Yet what is clear is that the problems accompanying these demands are currently underappreciated. This raises the question of who should be tasked with the increasing interpretation needs of data in the health care domain. Visions of data-rich medicine often imply that doctors should or will take on this work, as reflected in frequent calls for better genomic or data literacy for health care professionals. In the past decade, there have been numerous calls for more training in several of the domains mentioned above, such as ethical concerns surrounding the communication of genetic data and related health risks to patients [42], or counseling patients about the advantages and pitfalls of Web-based or commercial sources of health information [69]. Some, such as Celi et al, call for increased training of medical students and residents in order to "creat[e] a medical culture that is aware of and respectful of the importance and potential power of data for supporting and improving both practice and research may be the most important and ultimately effective element" [53]. At the moment, although health care professionals are seen as the first in line to take on this additional work, allowances are not made in schedules or training to accommodate meaningful engagement with the social complexities of data in medicine. Even if actors find the time to engage in the various types of data work, not all can acquire the necessary skills. Finally, many of the tasks described above take place outside health professionals' sphere of influence entirely.

Throughout this paper, we have proposed a few possible ways of addressing the emerging forms of data work identified here, ranging from the creation of a new profession dedicated to help both patients and providers assess and understand diverse kinds of health data, to greater involvement and creation of guidelines by scientific and academic societies, to raising expectations through regulatory frameworks for how mechanisms such as informed consent are operationalized across novel research practices. However, none of these approaches alone will be sufficient for taking on the myriad aspects of data work that we have outlined, as well as those that will continue to emerge in the future. Although the focus of this paper has been on the identification of the contours of the phenomenon we are calling *data work*, further attention is needed to analyze and consider other solutions for addressing these concerns. Importantly, some aspects of data work can neither be delegated to professionals nor addressed completely through better guidelines or greater public discourse. Hence, the current landscape of big data in medicine remains open for new proposals, such as how such work can or should be acknowledged or even reimbursed. What other tools—conceptual, analytic, instructive, or collaborative—would be helpful for navigating increasingly complex data use? What would be a fair division of work? What responsibilities should corporations using health data have, beyond compliance with data protection regulations? Our intent is that by making these forms of work more explicit and transparent, more appropriate ways of addressing data work can be devised in future.

## Conclusions

In addition to the established challenges surrounding data collection, storage, analysis, and security, pressing questions have arisen around: how to enable the appropriate use of technologies and engagement with health data outside of the structured environment of health care; what the utility, quality, and possibilities of data collected from wearable devices or smartphones will be for clinical practice; strategies to avoid the digital health divide; how to distinguish data noise from clinically actionable health resources for patients; how to contextualize health data gained through Web-based tests or digital technologies; and how to foster conversations surrounding the ethical concerns of big data between different stakeholders in health care and society. Of course, the various forms of work included within the categories of supporting digital tool use, contextualization, and inclusion and integration cannot be neatly disentangled. Conversations between different actors in the health care domain are necessary to determine what types of data and data use are feasible, ethical, and cost-effective in particular situations. Although we expect that AI applications such as deep learning will be of great help in matters such as the interpretation of data, the analysis above has shown that the task of interpretation is not something that can be devolved to machines entirely.

A critical thread that runs throughout the forms of data work identified here is that of context: data work does not involve questions of absolutes but rather of contingencies. What is relevant, important, or significant for one individual may not apply to the next. Data, just like the experience of health and illness, are profoundly dependent upon the social world in which they exist. As we have shown in this paper, the turn toward data-rich health care has created new forms of data work and expertise. Data work needs to be named and recognized as the human endeavors that make digital advances meaningful in medicine. We argue that greater attention is needed for the very craft of deriving choices, narratives, and practices from our data and that the current medical system is not equipped to take on this challenge alone. If the great potential of data-rich medicine to improve future clinical care is to be realized, the new data work that patients, health professionals, and other actors increasingly contribute must be recognized as an important and multifaceted task.

### Conflicts of Interest

None declared.

### References

1. Neff G. Why big data won't cure us. Big Data 2013 Sep;1(3):117-123 [FREE Full text] [doi: 10.1089/big.2013.0029] [Medline: 25161827]
2. Reddy MC, Gorman P, Bardram J. Special issue on supporting collaboration in healthcare settings: the role of informatics. Int J Med Inform 2011 Aug;80(8):541-543. [doi: 10.1016/j.ijmedinf.2011.05.001] [Medline: 21680235]

3. Aarts J. A sociotechnical perspective of health information technology. Int J Med Inform 2013 Dec;82(12):1133-1135. [doi: 10.1016/j.ijmedinf.2013.10.007] [Medline: 24216291]
4. Bjørn P, Kensing F. Special issue on information infrastructures for healthcare: the global and local relation. Int J Med Inform 2013 May;82(5):281-282. [doi: 10.1016/j.ijmedinf.2013.01.004] [Medline: 23422271]
5. Leonelli S. Data interpretation in the digital age. Perspect Sci 2014 Sep 12;22(3):397-417 [FREE Full text] [doi: 10.1162/POSC_a_00140] [Medline: 25729262]
6. Stein R. National Public Radio. 2018 Jun 18. Results Of At-Home Genetic Tests For Health Can Be Hard To Interpret URL: https://tinyurl.com/y974fodb [accessed 2018-06-25]
7. Kolata G, Murphy H. The New York Times. 2018 Apr 27. The Golden State Killer Is Tracked Through a Thicket of DNA, and Experts Shudder URL: https://tinyurl.com/y8yqd25c [accessed 2019-06-10]
8. Saey TH. Science News. 2019 Feb 6. What FamilyTreeDNA Sharing Genetic Data With Police Means For You URL: https://www.sciencenews.org/article/family-tree-dna-sharing-genetic-data-police-privacy [accessed 2019-06-10]
9. Sifferlin A. Time. 2017 Mar 16. Why Perfectly Healthy People Are Using Diabetes Monitors URL: http://time.com/4703099/continuous-glucose-monitor-blood-sugar-diabetes/ [accessed 2019-06-10]
10. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown; 2016.
11. Deutscher Ethikrat. 2017. [Big Data and Health - Data Sovereignty as Informational Freedom Design] URL: https://tinyurl.com/y6qnjhrb [accessed 2019-06-10]
12. Nuffield Council on Bioethics. 2014. The Collection, Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues URL: http://nuffieldbioethics.org/wp-content/uploads/Biodata-a-guide-to-the-report-PDF.pdf [accessed 2019-06-10]
13. Su P. Direct-to-consumer genetic testing: a comprehensive view. Yale J Biol Med 2013 Sep;86(3):359-365 [FREE Full text] [Medline: 24058310]
14. Richter G, Krawczak M, Lieb W, Wolff L, Schreiber S, Buyx A. Broad consent for health care–embedded biobanking: understanding and reasons to donate in a large patient sample. Genet Med 2018;20(1):76-82. [doi: 10.1038/gim.2017.82] [Medline: 28640237]
15. Whetton A. Institute of Infectious Disease and Molecular Medicine. 2019. Protein Biomarkers For Precision Medicine: Development of Large Scale Integrated Platform for Clinical Proteomics.
16. Neff G, Tanweer A, Fiore-Gartland B, Osburn L. Critique and contribute: a practice-based framework for improving critical data studies and data science. Big Data 2017;5(2):85-97 [FREE Full text] [doi: 10.1089/big.2016.0050] [Medline: 28632445]
17. Leonelli S. Data-Centric Biology: A Philosophical Study. Chicago: University Of Chicago Press; 2016.
18. Ebeling MF. Healthcare and Big Data: Digital Specters and Phantom Objects. New York: Palgrave Macmillan; 2016.
19. Tempini N, Leonelli S. Genomics and big data in biomedicine. In: Routledge Handbook of Genomics, Health and Society. UK: Routledge; 2018:44-51.
20. Strauss AL, Fagerhaugh S, Suczek B, Wiener C. The work of hospitalized patients. Soc Sci Med 1982;16(9):977-986. [doi: 10.1016/0277-9536(82)90366-5] [Medline: 7112176]
21. Strauss A. The articulation of project work: an organizational process. Sociol Q 1988;29(2):163-178. [doi: 10.1111/j.1533-8525.1988.tb01249.x]
22. Strauss AL. Social Organization of Medical Work. New Jersey: Transaction Publishers; 1997.
23. Clarke A, Mamo L, Fishman JR, Shim JK, Fosket JR. Biomedicalization: technoscientific transformations of health, illness, and US biomedicine. Am Sociol Rev 2003 Apr;68(2):161-194. [doi: 10.2307/1519765]
24. Stacey M. Who are the health workers? Patients and other unpaid workers in health care. Econ Ind Democr 1984;5(2):157-184. [doi: 10.1177/0143831X8452002]
25. Prainsack B. Personalized Medicine: Empowered Patients in the 21st Century?. New York: NYU Press; 2017.
26. Knoppers BM, Thorogood AM. Ethics and big data in health. Curr Opin Syst Biol 2017 Aug;4:53-57. [doi: 10.1016/j.coisb.2017.07.001]
27. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. Biomed Res Int 2015;2015:370194 [FREE Full text] [doi: 10.1155/2015/370194] [Medline: 26229957]
28. Loiselle CG, Ahmed S. Is connected health contributing to a healthier population? J Med Internet Res 2017;19(11):e386 [FREE Full text] [doi: 10.2196/jmir.8309] [Medline: 29127077]
29. Hesse BW, Greenberg AJ, Rutten LJ. The role of internet resources in clinical oncology: promises and challenges. Nat Rev Clin Oncol 2016 Dec;13(12):767-776. [doi: 10.1038/nrclinonc.2016.78] [Medline: 27273045]
30. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016 Apr 19;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
31. Klerings I, Weinhandl AS, Thaler KJ. Information overload in healthcare: too much of a good thing? Z Evid Fortbild Qual Gesundhwes 2015;109(4-5):285-290. [doi: 10.1016/j.zefq.2015.06.005] [Medline: 26354128]
32. Kung J, Wu CT. Leveling the playing field: closing the gap in public awareness of genetics between the well served and underserved. Hastings Cent Rep 2016;46(5):17-20. [doi: 10.1002/hast.613] [Medline: 27649825]

33. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
34. Del Savio L, Prainsack B, Buyx A. Motivations of participants in the citizen science of microbiomics: data from the British gut project. Genet Med 2017;19(8):959-961. [doi: 10.1038/gim.2016.208] [Medline: 28125088]
35. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the internet as a resource for health information: results from a California population-based study. Int J Med Inform 2017;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
36. Hong YA, Zhou Z, Fang Y, Shi L. The digital divide and health disparities in China: evidence from a national survey and policy implications. J Med Internet Res 2017;19(9):e317 [FREE Full text] [doi: 10.2196/jmir.7786] [Medline: 28893724]
37. Lorence DP, Park H, Fox S. Racial disparities in health information access: resilience of the digital divide. J Med Syst 2006 Aug;30(4):241-249. [doi: 10.1007/s10916-005-9003-y] [Medline: 16978003]
38. Fox S, Purcell K. Pew Research Center. 2010. Chronic Disease and the Internet URL: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2019-06-10]
39. Casper M, Moore LJ. Missing Bodies: The Politics Of Visibility. New York: NYU Press; 2009.
40. NHS. 2019 Feb. The Topol Review. Preparing the Healthcare Workforce to Deliver the Digital Future: An Independent Report on Behalf of the Secretary of State for Health and Social Care URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2019-06-10]
41. Gorman D, Kashner TM. Medical graduates, truthful and useful analytics with big data, and the art of persuasion. Acad Med 2018 Aug;93(8):1113-1116. [doi: 10.1097/ACM.0000000000002109] [Medline: 29280752]
42. Badalato L, Kalokairinou L, Borry P. Third party interpretation of raw genetic data: an ethical exploration. Eur J Hum Genet 2017;25(11):1189-1194 [FREE Full text] [doi: 10.1038/ejhg.2017.126] [Medline: 28832567]
43. Annes JP, Giovanni MA, Murray MF. Risks of presymptomatic direct-to-consumer genetic testing. N Engl J Med 2010 Sep 16;363(12):1100-1101. [doi: 10.1056/NEJMp1006029] [Medline: 20843242]
44. Fiske A, Buyx A, Prainsack B. Health information counselors: a new profession for the age of big data. Acad Med 2019 Jan;94(1):37-41 [FREE Full text] [doi: 10.1097/ACM.0000000000002395] [Medline: 30095453]
45. Granado-Font E, Flores-Mateo G, Sorlí-Aguilar M, Montaña-Carreras X, Ferre-Grau C, Barrera-Uriarte ML, OBSBIT Study Group. Effectiveness of a smartphone application and wearable device for weight loss in overweight or obese primary care patients: protocol for a randomised controlled trial. BMC Public Health 2015 Jun 4;15:531 [FREE Full text] [doi: 10.1186/s12889-015-1845-8] [Medline: 26041131]
46. Lewis Jr GK, Langer MD, Henderson Jr CR, Ortiz R. Design and evaluation of a wearable self-applied therapeutic ultrasound device for chronic myofascial pain. Ultrasound Med Biol 2013 Aug;39(8):1429-1439. [doi: 10.1016/j.ultrasmedbio.2013.03.007] [Medline: 23743101]
47. Mummah S, Robinson TN, Mathur M, Farzinkhou S, Sutton S, Gardner CD. Effect of a mobile app intervention on vegetable consumption in overweight adults: a randomized controlled trial. Int J Behav Nutr Phys Act 2017;14(1):125 [FREE Full text] [doi: 10.1186/s12966-017-0563-2] [Medline: 28915825]
48. Macdonald GG, Townsend AF, Adam P, Li LC, Kerr S, McDonald M, et al. eHealth technologies, multimorbidity, and the office visit: qualitative interview study on the perspectives of physicians and nurses. J Med Internet Res 2018;20(1):e31 [FREE Full text] [doi: 10.2196/jmir.8983] [Medline: 29374004]
49. Anstey Watkins J, Goudge J, Gómez-Olivé FX, Huxley C, Dodd K, Griffiths F. mHealth text and voice communication for monitoring people with chronic diseases in low-resource settings: a realist review. BMJ Glob Health 2018;3(2):e000543 [FREE Full text] [doi: 10.1136/bmjgh-2017-000543] [Medline: 29527356]
50. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]
51. Turrini M, Prainsack B. Beyond clinical utility: the multiple values of DTC genetics. Appl Transl Genom 2016;8:4-8 [FREE Full text] [doi: 10.1016/j.atg.2016.01.008] [Medline: 27047753]
52. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
53. Celi LA, Davidzon G, Johnson AE, Komorowski M, Marshall DC, Nair SS, et al. Bridging the health data divide. J Med Internet Res 2016 Dec 20;18(12):e325 [FREE Full text] [doi: 10.2196/jmir.6400] [Medline: 27998877]
54. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013 Oct;22(Suppl 2):ii21-ii27 [FREE Full text] [doi: 10.1136/bmjqs-2012-001615] [Medline: 23771902]
55. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
56. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
57. Fernandez-Luque L, Karlsen R, Bonander J. Review of extracting information from the social web for health personalization. J Med Internet Res 2011 Jan 28;13(1):e15 [FREE Full text] [doi: 10.2196/jmir.1432] [Medline: 21278049]

58. Syed-Abdul S, Fernandez-Luque L, Jian WS, Li YC, Crain S, Hsu MH, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013 Feb 13;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
59. Fiske A, Prainsack B, Buyx A. Survey of Health Care Practitioner Assessments of Self Care in Schleswig-Holstein, Germany 2019.
60. Aitken M, Lyle J. IQVIA. 2015 Sep. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf
61. The European Commission. 2015. The Ethical Implications of New Health Technologies and Citizen Participation URL: http://ec.europa.eu/research/ege/pdf/opinion-29_ege.pdf [accessed 2019-06-10]
62. Pasquale F. The Black Box Society: The Secret Algorithms That Control Money And Information. London, England: Harvard University Press; 2016.
63. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht UV. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
64. Harris A, Kelly S, Wyatt S. CyberGenetics: Health Genetics and New Media. UK: Routledge; 2016.
65. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People-an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach (Dordr) 2018;28(4):689-707 [FREE Full text] [doi: 10.1007/s11023-018-9482-5] [Medline: 30930541]
66. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019 May 9;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
67. Ruckenstein M, Schüll ND. The datafication of health. Annu Rev Anthropol 2017;46:261-278. [doi: 10.1146/annurev-anthro-102116-041244]
68. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, Kanellopoulou N, et al. From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet 2012 Apr 3;13(5):371-376 [FREE Full text] [doi: 10.1038/nrg3218] [Medline: 22473380]
69. Nuffield Council on Bioethics. 2010. Medical Profiling and Online Medicine: The Ethics of 'Personalised Healthcare' in a Consumer Age URL: https://tinyurl.com/zpq4pwb [accessed 2019-06-10]
70. Capper D, Jones DT, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature 2018;555(7697):469-474 [FREE Full text] [doi: 10.1038/nature26000] [Medline: 29539639]
71. Klingmüller U. Ethics, bio-politics and regulation of precision medicine. Hamburg; 2018 Mar 28 Presented at: International Cluster Symposium: Precision Medicine in Chronic Inflammation; March 26-27, 2018; Hamburg, Germany URL: https://tinyurl.com/yxh7thkv
72. Schueller SM, Washburn JJ, Price M. Exploring mental health providers' interest in using web and mobile-based tools in their practices. Internet Interv 2016 May;4(2):145-151 [FREE Full text] [doi: 10.1016/j.invent.2016.06.004] [Medline: 28090438]
73. Beauchamp TL, Childress JF. Principles Of Biomedical Ethics. Oxford, USA: Oxford University Press; 2012.
74. Rumbold JM, Pierscionek B. The effect of the general data protection regulation on medical research. J Med Internet Res 2017;19(2):e47 [FREE Full text] [doi: 10.2196/jmir.7108] [Medline: 28235748]
75. Thompson B. Welcome Trust. 2016 Jul. Analysis: Research and the General Data Protection Regulation URL: https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf [accessed 2019-06-10]
76. Dreyer NA, Blackburn S, Hliva V, Mt-Isa S, Richardson J, Jamry-Dziurla A, et al. Balancing the interests of patient data protection and medication safety monitoring in a public-private partnership. JMIR Med Inform 2015 Apr 15;3(2):e18 [FREE Full text] [doi: 10.2196/medinform.3937] [Medline: 25881627]
77. US Food and Drug Administration. 2015. Mobile Medical Applications URL: https://www.fda.gov/medical-devices/digital-health/mobile-medical-applications [accessed 2019-06-10]
78. LegiTech. 2017 Mar 30. GDPR Series: Part 8 - Leeway Granted to Member State National and Supervisory Authorities URL: https://tinyurl.com/y5gny58v [accessed 2019-04-23]
79. Snell E. Health IT Security. 2018 Jan 22. How Compliance, Data Security Needs Shift with Big Data Push URL: https://healthitsecurity.com/news/how-compliance-data-security-needs-shift-with-big-data-push [accessed 2019-04-23]
80. McMahon A, Buyx A, Prainsack B. Big data governance needs more collective responsibility: the role of harm mitigation in the governance of data use in medicine and beyond. Med Law Rev 2019 [FREE Full text]
81. Sanderson SC, Brothers KB, Mercaldo ND, Clayton EW, Antommaria AH, Aufox SA, et al. Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US. Am J Hum Genet 2017 Mar 2;100(3):414-427 [FREE Full text] [doi: 10.1016/j.ajhg.2017.01.021] [Medline: 28190457]
82. Kayyali R, Hesso I, Ejiko E, Gebara SN. A qualitative study of telehealth patient information leaflets (TILs): are we giving patients enough information? BMC Health Serv Res 2017;17(1):362 [FREE Full text] [doi: 10.1186/s12913-017-2257-5] [Medline: 28526026]
83. Prainsack B, Buyx A. Solidarity in Biomedicine and Beyond. Cambridge: Cambridge University Press; 2017.

84. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: a qualitative study. J Med Internet Res 2016 Apr 15;18(4):e66 [FREE Full text] [doi: 10.2196/jmir.5011] [Medline: 27083521]
85. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient access to electronic health records: differences across ten countries. Health Policy Technol 2017;7(1):44-56. [doi: 10.1016/j.hlpt.2017.11.003]
86. Wass S, Vimarlund V, Ros A. Exploring patients' perceptions of accessing electronic health records: innovation in healthcare. Health Informatics J 2019 Mar;25(1):203-215. [doi: 10.1177/1460458217704258] [Medline: 28457195]
87. Pormeister K. Genetic data and the research exemption: is the GDPR going too far? Int Data Priv Law 2017 May;7(2):137-146. [doi: 10.1093/idpl/ipx006]

## Abbreviations

**AI:** artificial intelligence
**EU:** European Union
**GDPR:** General Data Protection Regulation
**HIC:** health information counselor
**HIPAA:** Health Insurance Portability and Accountability Act

*Edited by G Eysenbach; submitted 25.07.18; peer-reviewed by C Shivade, X Garcia-Eroles; comments to author 01.02.19; revised version received 27.03.19; accepted 26.04.19; published 09.07.19.*

*Please cite as:*
*Fiske A, Prainsack B, Buyx A*
*Data Work: Meaning-Making in the Era of Data-Rich Medicine*
*J Med Internet Res 2019;21(7):e11672*
*URL: https://www.jmir.org/2019/7/e11672/*
*doi:10.2196/11672*
*PMID:31290397*

Review

# Internet-Based Interventions for Carers of Individuals With Psychiatric Disorders, Neurological Disorders, or Brain Injuries: Systematic Review

Lucy Spencer[1], MA (Oxon), MSc; Rachel Potterton[1], BA, MSc; Karina Allen[1,2], BA (Hons), MPsych, PhD; Peter Musiat[1], PhD; Ulrike Schmidt[1,2], MD, PhD, FRCPsych

[1]Section of Eating Disorders, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom

[2]The Eating Disorders Service, Maudsley Hospital, South London & Maudsley National Health Service Foundation Trust, London, United Kingdom

**Corresponding Author:**
Lucy Spencer, MA (Oxon), MSc
Section of Eating Disorders
Institute of Psychiatry, Psychology, and Neuroscience
King's College London
16 De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 2078485608
Email: lucy.spencer@kcl.ac.uk

## *Abstract*

**Background:** Nonprofessional carers who provide support to an individual with a psychiatric or neurological disorder will often themselves experience symptoms of stress, anxiety, or low mood, and they perceive that they receive little support. Internet-based interventions have previously been found to be effective in the prevention and treatment of a range of mental health difficulties in carers.

**Objective:** This review seeks to establish the status of internet-based interventions for informal (nonprofessional) carers of people with psychiatric or neurological disorders by investigating (1) the number and quality of studies evaluating the efficacy or effectiveness of internet-based carer interventions and (2) the impact that such interventions have on carer mental health, as well as (3) how internet-based interventions compare with other intervention types (eg, face-to-face treatment).

**Methods:** A systematic literature search was conducted in January 2019 using the EMBASE (1974-present), Ovid MEDLINE (1946-present), PsychARTICLES, PsychINFO (1806-present), and Global Health (1973-present) databases, via the Ovid Technologies database. Search terms included carer, caregiver, online, technology, internet-based, internet, interactive, intervention, and evaluation. Studies selected for inclusion in this review met the following predetermined criteria: (1) delivering an intervention aimed primarily at informal carers, (2) carers supporting individuals with psychiatric disorders, stroke, dementia, or brain injury, (3) the intervention delivered to the carers was primarily internet based, (4) the study reported a pre- and postquantitative measure of carer depression, anxiety, stress, burden, or quality of life, (5) appeared in a peer-reviewed journal, and (6) was accessible in English.

**Results:** A total of 46 studies were identified for inclusion through the detailed search strategy. The search was conducted, and data were extracted independently by 2 researchers. The majority of studies reported that 1 or more measures relating to carer mental health improved following receipt of a relevant intervention, with interventions for carers of people with traumatic brain injury showing a consistent link with improved outcomes.

**Conclusions:** Studies investigating internet-based interventions for carers of individuals with diverse psychiatric or neurological difficulties show some evidence in support of the effectiveness of these interventions. In addition, such interventions are acceptable to carers. Available evidence is of varying quality, and more high-quality trials are needed. Further research should also establish how specific intervention components, such as structure or interactivity, contribute to their overall efficacy with regard to carer mental health.

***(J Med Internet Res 2019;21(7):e10876)*** doi:10.2196/10876

**KEYWORDS**

internet; carers; mental health; technology; review

# Introduction

## Background

Nonprofessional (or informal) carers are individuals who provide free-of-charge care for another person (usually a family member or friend), who would find it difficult to cope without the carer's support. Informal carers play a crucial role in providing both practical and emotional care for individuals with a wide range of difficulties, including physical and mental health difficulties, disabilities, or addictions. It is estimated that during 2015, 6.8 million people in the United Kingdom provided unpaid care to a close other, a 16.5% increase from 2001, reflecting an economic value of £132 billion of informal care per year, almost double the value in 2001 [1]. In addition, the support of informal carers may reduce and delay hospital admissions [2], thus further reducing the burden on health care systems.

It has been widely documented that caring for a loved one with a long-term illness can have a multitude of effects on the informal carer, including increased levels of perceived burden [3], feelings of entrapment, shame, guilt [4], and higher rates of physical symptoms, such as fatigue, headaches, and weight loss [5]. It is important to recognize that disorder- or patient-related factors can profoundly impact carers' experience of caregiving. Such factors may include the nature and severity of different symptoms, societal reactions, and certain pathologies, which may vary widely both within and between disorders. For example, when comparing carers of people with schizophrenia and carers of people with long-term physical disorders, levels of subjective burden were found to be higher in carers of people with schizophrenia and brain diseases than in other groups [6]. Furthermore, levels of social support available to carers of individuals with schizophrenia were found to be lower than for carers of individuals with physical disorders.

In addition to the differential impact that various diagnoses and symptoms can have on carer difficulties, it is also important to consider how individual differences in perceptions of burden affect the experience of caring. Caregiver Identity Theory [7] posits that the main source of carer distress is identity discrepancy—the disparity between the activities they are required to carry out as a carer and their own views of self (or *identity standard*). This may explain why there is a wide variation in perceived burden or distress in carers who, on the face of it, are required to carry out similar caring responsibilities (eg, carers of people with dementia), and it may be a helpful dimension to explore when considering possible interventions or support plans. A range of carer-focused face-to-face interventions, often delivered in group formats, have been found to reduce psychological distress and improve the quality of life of individuals caring for people with severe mental health difficulties [8]. However, some carers may find it difficult to attend regular appointments because of time constraints, or they may have concerns regarding privacy or stigma [9]. Internet-based interventions have been found to be effective in prevention and treatment of a range of psychiatric disorders, including depression [10], anxiety [11], posttraumatic stress disorder [12], and eating disorders [13]. As carers often experience elevated levels of depression and stress, as well as reduced general well-being [14], internet interventions to improve carers' own mental health should be considered as a potentially viable option. Previous reviews of the impact of internet-based interventions on carer distress have focused on carers of people with a broad range of mental and physical disorders (including dementia, cancer, mental health difficulties, and hip fractures), and they have reported positive or mixed findings [15,16]. In contrast, this review focuses specifically on carers of people with psychiatric disorders, neurological disorders (dementia, stroke), or brain injury. We decided to restrict our inclusion criteria to carers of individuals with these disorders specifically, as the burden of caring for someone whose primary difficulty relates to his or her cognitive abilities or mental health may be very different from that of caring for someone whose condition primarily impacts on his or her physical health (eg, cancer). Some studies focus on whether internet-based interventions can help increase carers' knowledge of their loved one's disorder [17] or teach them relevant skills to manage or change their loved one's behavior—for example, children with attention-deficit hyperactivity disorder (ADHD) [18].

## Objectives

Although improving carer knowledge and skills is important, this review focuses on whether internet-based interventions can improve carer mental health, and this review has therefore only included studies that measure aspects of this, such as depression, anxiety, stress, burden, or perceived quality of life. Specifically, this review seeks to establish the status of internet-based interventions for informal carers of people with psychiatric or neurological disorders by investigating (1) the number and quality of studies evaluating the efficacy or effectiveness of internet-based carer interventions and (2) the impact that such interventions have on carer mental health, (3) as well as how internet-based interventions compare with other intervention types (eg, face-to-face treatment).

# Methods

## Eligibility Criteria

The papers selected for inclusion in this review met the following predetermined criteria: (1) delivering an intervention aimed primarily at informal (nonprofessional) carers of (2) children or adults with psychiatric disorders, stroke, dementia, or brain injury; (3) the intervention delivered to the carers was primarily internet based, (4) the study reported a pre- and postquantitative measure of carer depression, anxiety, stress, burden, or quality of life, (5) appeared in a peer-reviewed journal, and (6) was accessible in English. Studies were excluded if the intervention was aimed exclusively at the patient rather than the carer. Papers exclusively reporting other measures of intervention efficacy (eg, increase in carer knowledge) were also excluded from this review.

### Information Sources

A systematic literature search was conducted in January 2019 using the EMBASE (1974-present), Ovid MEDLINE (1946-present), PsychARTICLES, PsychINFO (1806-present), and Global Health (1973-present) databases, via the Ovid Technologies database. Searches of reference lists of articles listed in this review, as well as relevant review papers, were also conducted. A total of 5 papers identified in the literature search, which could not be obtained, were requested from their authors via ResearchGate. Of these, 2 authors responded by sending the full text of their study to be assessed for eligibility. The search was limited to papers that could be accessed in English.

### Search Strategy

Search terms were the following; (carer OR caregiver OR care-giver OR carers) AND (online OR on-line OR technology OR internet-based OR interactive OR internet) AND (intervention OR evaluation). This search strategy returned 46 studies that met each of the inclusion criteria detailed above.

### Data Collection Process

A data extraction sheet (adapted from the Cochrane Consumers and Communication Review Group's data extraction template) was developed and pilot tested on 5 randomly selected studies to be included in the review and refined accordingly following the pilot testing. Data were extracted from the included studies by 1 study author (LS) and checked by a second author (RP). Disagreements between reviewers were resolved by consensus. No authors were contacted for further information.

### Data Items

For each study within the review, we extracted participant characteristics (number per study arm, average age, and gender), details of the internet-based intervention (including intervention name, content, average number of sessions, and duration), and details of the control group where applicable. Regarding study findings, we extracted data regarding the statistical significance of quantitative findings of carer psychological health and the time points at which outcome data were collected. In addition, we identified the primary outcome(s) where this was specified and extra data regarding any qualitative findings with particular relevance to carer outcomes and mental health.

### Rating Evidence of Intervention Effectiveness

Each of the studies included in the review was rated on the effectiveness of the intervention employed, in terms of the extent to which the intervention had a statistically significant impact on outcomes relating to carer depression, anxiety, stress, burden, or quality of life. Studies were given 1 of 3 ratings:

- Intervention shows *clear association* with positive outcomes relating to carer depression, anxiety, stress, burden, or quality of life (half or more outcome measures show statistically significant and positive impact of intervention). *Effectiveness of intervention score=3.*
- Intervention shows *some association* with positive outcomes relating to carer depression, anxiety, stress, burden, or quality of life (fewer than half, but at least one outcome measure show(s) statistically significant and positive impact of intervention). *Effectiveness of intervention score=2.*
- Intervention shows *very little or no association* with positive outcomes relating to carer distress, anxiety, stress, burden, or quality of life (no outcomes showing a statistically significant and positive impact of intervention). *Effectiveness of intervention score=1.*

### Risk of Bias in Individual Studies

Risk of bias within individual studies was rated independently by 2 researchers (LS and RP). Studies were assessed using a scale developed previously to evaluate risk of bias and study quality in a review of interventions for individuals with anorexia nervosa (AN) [19]. A table displaying the risk of bias of each individual study can be found in Multimedia Appendix 1. Randomized controlled trials (RCTs) were given a rating out of 20 (the total number of items). RCTs receiving a score of 14 or over were deemed to be *high quality*, RCTs receiving a score of above 8 and below 14 were deemed to be of *moderate quality*, and RCTs studies receiving a score of 8 or below were deemed to be *low quality*. Studies that did not employ an RCT design were rated on the 13 items of the scale relevant to non-RCTs. Of these non-RCT studies, those receiving a score of 10 or over were deemed to be *high quality*, those receiving a score of above 6 and below 10 were deemed to be of *moderate quality*, and those receiving a score of 6 or below were deemed to be *low quality*. Disagreements were resolved by discussion. Of the 46 studies identified for inclusion in the review, 16 RCTs and 0 non-RCTs were rated as high quality, 11 RCTs and 6 non-RCTs were rated as moderate quality, and 2 RCTs and 11 non-RCTs were rated as low quality.

### Study Selection

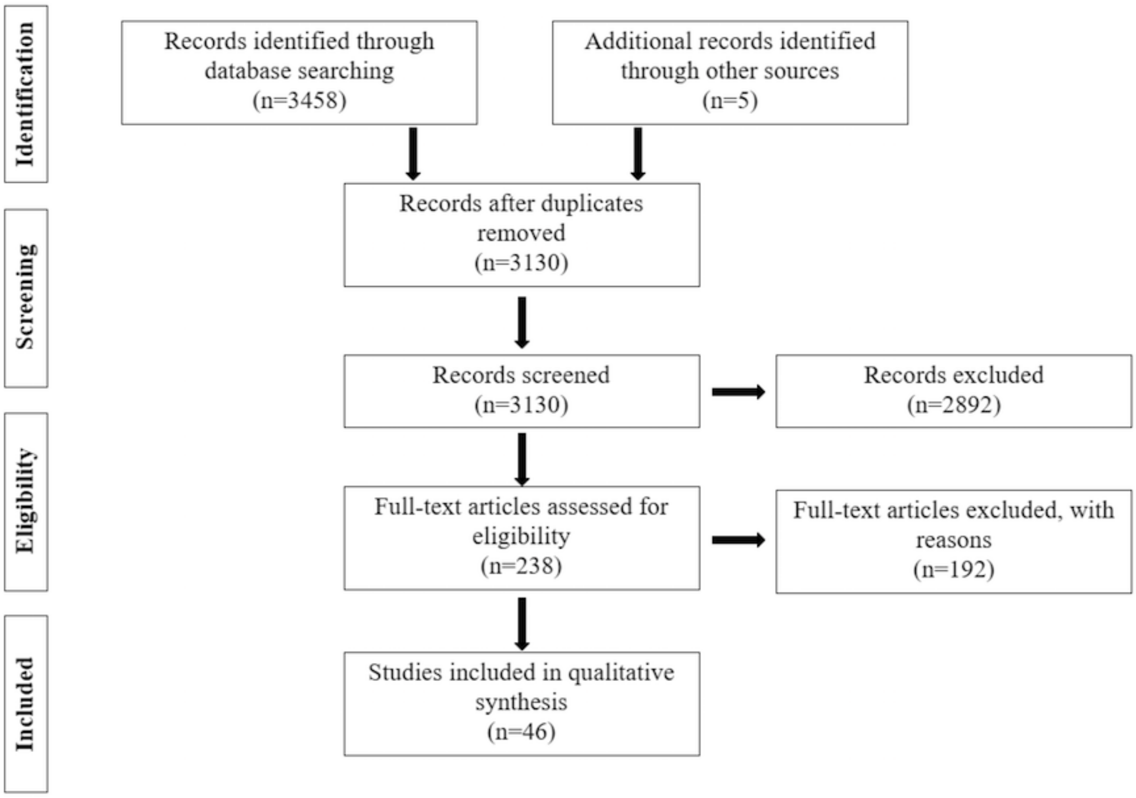
Eligibility assessment for each study to be included in the systematic review was performed independently by 2 reviewers (LS and RP), using the eligibility criteria detailed above. Disagreements between reviewers were resolved by consensus. A search of EMBASE, Ovid MEDLINE, PsychARTICLES, PsychINFO, and Global Health returned a total of 3458 studies. A total of 5 additional citations were identified through searching the reference lists of relevant papers. After a preliminary screening, the full text of 238 articles were examined in detail, of which 192 were excluded. Reasons for the exclusion of the studies examined in detail can be found in Table 1.

Agreement between the 2 reviewers during the selection of abstracts, as measured by Cohen kappa, was 0.860, and agreement during selection of full texts for inclusion was 0.892, both of which are regarded as excellent. A total of 46 studies met inclusion criteria and were included in the review. A flow diagram detailing the selection of studies for inclusion can be found in Figure 1.

**Table 1.** Reasons for exclusion of studies examined in detail (n=192).

| Reason for excluding study | Studies excluded, n (%) |
|---|---|
| Review article | 27 (14.1) |
| Does not report an intervention | 10 (5.2) |
| Qualitative data only | 16 (8.3) |
| Carer recipient not suffering from mental health, dementia, stroke, or brain injury | 22 (11.5) |
| Unable to access | 12 (6.25) |
| Not available in English | 4 (2.1) |
| Not an article in a peer-reviewed journal | 28 (14.6) |
| Does not report pre/post measure of burden, stress, depression, anxiety, or quality of life | 25 (13.0) |
| Study protocol | 11 (5.7) |
| Does not report an internet-based intervention | 29 (15.1) |
| Intervention aimed at sufferer rather than carer | 6 (3.1) |
| Professional caregivers | 2 (1.0) |

**Figure 1.** Flowchart of study selection for the review.




## Results

### Study Quality

Of the 46 studies identified for inclusion in the review, 16 RCTs and 0 non-RCTs were rated as high quality, 11 RCTs and 6 non-RCTs were rated as moderate quality, and 2 RCTs and 11 non-RCTs were rated as low quality.

### Psychiatric Disorders

Studies regarding carers of individuals with psychiatric disorders are summarized in Multimedia Appendix 2. A total of 12 such studies were identified [18,20-30]. For the purpose of this review, studies are grouped and presented below by specific disorder (AN, Schizophrenia or Schizoaffective disorder, ADHD, and studies describing mixed mental health difficulties).

#### *Anorexia Nervosa*

A total of 2 RCTs explored the efficacy of a sequential, 8-modular (with participants completing approximately 1 module per week) internet-based intervention (Overcoming Anorexia Online; OAO), based upon a systemic, cognitive behavioral therapy (CBT) framework, in supporting carers of adults with AN. In a study rated to be of high quality, in a comparison of carers receiving OAO to those receiving support as usual from the charity B-eat, OAO was found to reduce levels of carer anxiety and depression to a greater extent than support as usual [20]. In a later study, rated to be of moderate quality, the OAO intervention with clinician guidance was compared with OAO without additional guidance [21]. No significant improvements were found in either group over time, regarding carer anxiety, depression, or stress. Improvements over time were observed across other measures (including negative experiences of caregiving and intrusiveness), although some of these improvements were observed in the group receiving clinician guidance, and some were observed in the group that did not receive guidance.

### *Depression*

In a high-quality RCT, participation in a self-management intervention, "E-care for Caregivers", for a period of 6 weeks [22] was not found to be associated with a decrease in carer distress. However, the intervention was rated as user friendly by participants, indicating that the further development and implementation of internet-based interventions may be acceptable to carers of individuals with depression.

### *Schizophrenia and Schizoaffective Disorder*

The efficacy of a telehealth intervention ("The Schizophrenia Guide") of 3 months duration, for individuals with schizophrenia (or schizoaffective disorder) and their carers, was compared within an RCT with care as usual, in a study considered to be of moderate quality. No between-group differences were found with regard to levels of perceived carer stress. However, individuals with schizophrenia allocated to the intervention group were found to display reduced stress levels [23]. Another study investigated the efficacy of a internet-based, multifamily intervention [24], compared with support as usual in a quasi-experimental trial, with the interventions delivered over a period of 12 months. This study also did not find a significant difference in levels of carer distress between intervention and control groups. However, family relationship stress improved over time in the intervention group, and the majority of users expressed high levels of satisfaction with the internet-based intervention.

### *Attention-Deficit Hyperactivity Disorder*

A small case series (n=8) investigating the impact of an 8-session, psychoeducational parenting program, delivered via videoconferencing, found improvements in both parent distress and child behavior over time [18]. In a recent study, 6 sessions of behavior training delivered over an intervention period of 25 weeks via videoconferencing technology was compared with the same intervention content, delivered in person to participants [25]. Carers allocated to the videoconferencing group did not report improvements in their own mental health, whereas those in the in-person group did. Families in both conditions reported similar levels of improvement in their child's level of functioning and comparable satisfaction.

### *Mixed Mental Health Difficulties*

Stjernswärd and Hansson [26], in a high quality RCT, compared a 10-week internet-based mindfulness program with a wait-list control, in adult carers supporting people with a wide range of diagnoses (including depression, anxiety disorders, psychosis, and autism spectrum disorders), and they found that, in addition to an improvement in mindfulness, those in the intervention group also reported a reduction in perceived stress and some aspects of quality of life and carer burden. In addition, the same authors reported results of a prepost comparison, in which participants completed an 8-week internet-based mindfulness program [27]. Similarly to the results of their RCT, improvements were found in carer quality of life and burden, perceived stress, and mindfulness. These improvements were largely maintained at 3-month follow-up. In their most recent study, these authors reported the effectiveness of a similar internet-based mindfulness program on a large number (n=398) of carers of people with mental or somatic illnesses [28]. Improvements in carer stress were again found in the intervention group at 8 weeks, maintained at follow-up. Within the experimental group, burden was found to decrease from pre to follow-up on both the objective and subjective subscales. A recent prepost comparison study [29] targeting carers of adolescents with mental health problems found that a 3-month intervention powered by a Moderated Online Social Therapy software platform was acceptable and safe for use by participants, and it found that after engaging with the program, participants showed a significant reduction on a measure of stress, although other measures relating to carer mental well-being (depression, anxiety, and psychological well-being) had not changed significantly by the end of the intervention. Finally, a recent RCT (considered to be of high quality) specifically focused on the needs of young carers (aged 16-25 years) of individuals with mental illness. This study compared the effectiveness of a internet-based intervention with "folder support" (participants in this condition were provided with a folder, containing information on available support services). Results regarding the efficacy of the internet-based intervention were mixed [30]. No between-group differences were observed with regard to carer stress, and although both groups displayed an increase in well-being, only the folder group displayed improvements in carer self-efficacy and quality of life.

## Stroke

A total of 4 studies (summarized in Multimedia Appendix 3) were identified, which tested internet-based interventions for carers of stroke survivors [31-34]. A total of 2 of these were RCTs, and they were rated as being of moderate and high quality, respectively [32,33], with the other 2 studies [31,34], comprising prepost comparisons, rated to be of low quality. Ranging in duration from 4 weeks [31] to 12 months [32,34], each of the 4 interventions provided relevant informative resources, in addition to contact with fellow carers and professionals through a range of channels, including email, message boards, internet-based chats, telephone, and videophone. With regard to carer psychological functioning, 2 of the 4 studies found no significant change in measures of depression, life satisfaction [32], burden, or mental health [34] in participants receiving the respective internet-based interventions. However, Smith et al [33] found a significant improvement in reported depression (including a clinically meaningful change) in the intervention group over time, whereas Graf et al [31] reported a decrease in both depressive symptoms and burden, regardless of total number of years spent caring. There appeared to be a relationship between intervention duration and carer outcomes—the 2 studies finding significant improvements over time [31,33] lasted 4 and 11 weeks, respectively, whereas the 2 of longer duration (12 months) [32,34] did not find any significant change in carer outcomes.

### Dementia

We identified a total of 22 studies documenting internet interventions for carers of people with dementia. A total of 12 of these included a comparison group (Multimedia Appendix 4; [17,35-45]), and 10 studies were conducted without a control group (Multimedia Appendix 5; [34,46-54]). Of the studies featuring a control group, 5 were rated as high quality, 4 were rated as moderate quality, and 3 were rated as low quality. Of the studies without a control group, 4 were rated as moderate quality, and 6 were rated as low quality. Intervention duration varied from 30 days [35] to 12 months [34,36,37].

Again, these studies reported mixed findings with regard to the impact of the interventions on measures of carer mental health, with 8 of the interventions showing a clear positive impact on carer outcomes (effectiveness score of 3), 5 of the interventions showing some positive impact on carer outcomes (effectiveness score of 2), and 9 found to have little or no positive impact on carer outcomes (effectiveness score of 1). Interventions reporting clear positive findings tended to be those of shorter duration, with 4 of the 8 lasting 9 weeks or less [35,46-48], in comparison to interventions reporting little or no positive impact, of which only 2 of the 9 studies lasted 9 weeks or less [38,49].

### Traumatic Brain Injury

This review found 9 studies [55-63] meeting the inclusion criteria for carers of people with traumatic brain injury (TBI; Multimedia Appendix 6). Of these studies, 8 featured interventions aimed specifically at carers of children and adolescents with TBI. All of these featured self-guided, modular interventions, followed by either a videoconference or Skype session with a therapist. Interventions varied in duration from 10 days [55] to 6 months [56-60], although a number of studies did not specify the length of the intervention [61-63]. Findings with regard to the effectiveness of these interventions were positive, with only 1 of the 9 studies failing to find an association between the intervention and positive carer outcomes [55].

## Discussion

### Principal Findings

This systematic review identified 46 studies investigating internet-based interventions for carers of individuals with psychiatric disorders, stroke, dementia, or TBI, with regard to their efficacy or effectiveness in improving or maintaining a range of facets of carer mental health. Findings for each of the different disorders are discussed below.

### Psychiatric Disorders

A total of 12 studies investigated the impact of an internet-based intervention on carers of individuals with a range of psychiatric disorders, including AN, depression, schizophrenia, ADHD, and mixed mental health difficulties. Individuals with these different disorders present with varying needs and support requirements, thus presenting different challenges for their carers, making it hard to compare across disorders. Moreover, study and intervention design varied considerably across studies, both of which potentially affect outcomes. Aspects of an intervention that may influence the impact of a specific intervention include the following: whether the intervention is theory-driven and based on a specific model of carer distress or, alternatively, how else the intervention content and format were decided upon (eg, focus groups, literature reviews, or clinician and researcher opinion), whether the intervention is delivered via a website or other internet-based technology (such as videoconferencing), whether the intervention is modular and sequential or allows unstructured exploration of a website, the extent to which the intervention is interactive (eg, is feedback given to participants on their knowledge, symptoms, or any other characteristics), whether and how it is supported, (eg, can participants interact with one another or communicate with clinicians and researchers), and, in addition, does the intervention contain elements other than text, such as video or audio features. Given the small number of studies available, there is a clear need for additional research. Possible future research directions are considered for each disorder below.

#### *Anorexia Nervosa*

Both studies concerning carers of individuals with AN tested aspects of the same intervention (OAO), which was derived from a model of carer distress, coproduced with carers, experts by experience, and professionals within the field [64]. In the first of these, the internet-based intervention was delivered with support from experienced clinicians, and improved carer distress was compared with usual support [20]. In the second study, which was small and underpowered, the addition of limited support by a trainee psychologist had no advantage over a internet-based intervention alone [21]. Taken together, these findings demonstrate that OAO can provide benefits for carer mental health, over and above the support typically offered to carers of individuals with AN. However, the impact of additional clinician support is currently unclear. This may be partly because of factors pertaining to the study methodology, including small sample size (n=37) and guidance being provided by trainee clinicians with limited experience in the field of eating disorders. Furthermore, the follow-up period (3-months) may be too short for carers to have fully honed and applied the skills taught within the program, given the chronic nature of AN, a carer's behavior patterns may be even more long standing and ingrained than those found in other disorders. Future research in this area should seek to establish the additional benefit (if any) of clinician support and establish whether any observed differences were maintained at longer-term follow-up. In addition, it would also be useful to extend this work to carers of individuals diagnosed with other eating disorders, such as bulimia nervosa or binge eating disorder.

#### *Depression*

Only 1 study [22] assessed the efficacy of an internet-based intervention for carers of people with depression. The intervention was interactive and modular and based on psychoeducation and CBT techniques. Despite high reported levels of user friendliness, the intervention was ineffective. Of note, over 50% of participating carers kept their study participation a secret from those they cared for. Possible interpretations of this include carers not wishing it to be known that they require support, either as they fear the person they support may feel upset or guilty or as they wish to appear strong toward their loved one; carers not thinking it relevant or useful to share their participation with their loved one; carers not being sufficiently engaged with the intervention to treat it as an important program worthy of sharing and discussion. Research on the reasons carers do or do not share their involvement with internet-based support programs may be beneficial. Regardless of the reason(s), it is possible that this degree of secrecy contributed to the lack of efficacy of the intervention, as it may have made it harder for participants to alter their behavior and apply skills learned via the internet-based package (without the person they support finding out that they have accessed support), and it may thus inadvertently maintain the illness and carers' own distress. Future interventions for carers of people with depression may need to address and remedy these issues.

### *Schizophrenia and Schizoaffective Disorder*

A total of 2 studies focused on carers of individuals with schizophrenia; 1 was a small underpowered RCT of moderate quality [23], and the other employed a quasi-experimental design [24], rated to be of moderate quality.

The intervention reported in the Rotondi et al's study [23] comprised a website, with the content aimed at both the individual with schizophrenia and the individual's carer. Although this is an interesting idea, it would be important to establish whether an intervention based on an evidence-based model of distress specifically targeted at carers is more effective than one targeting both the individual with schizophrenia and the individual's carer, as these 2 groups may have different needs. Furthermore, although both interventions permitted interaction with fellow participants, neither employed a modular approach. Research in other mental health populations found modular treatment to result in better outcomes compared with standard treatment [65]. Thus, it may be helpful for future research in this area to establish whether a more structured approach may be also associated with improvements in carer mental health. In addition, it may be helpful to establish whether the observed improvements in family relations may lead to improvements in carer mental health, if they are maintained over a longer period of time.

### *Attention-Deficit Hyperactivity Disorder*

Neither of the 2 studies of interventions for carers of individuals with ADHD were RCTs; in a small case series [18], the group Triple P Parenting Program showed promise in reducing carer distress. A second study [25] involved a subset of data drawn from a larger study comparing the same intervention face to face or delivered via teletherapy. Carer distress only improved in the face-to-face delivery group. Although these findings are promising, RCTs are needed to further elucidate the role of internet-based interventions in supporting carers of people with ADHD. Qualitative data from carers may help explore carers' views on the relative merits of face-to-face versus internet-based interventions, as in the Tse et al study [25], only carers receiving in-person training displayed improvements in their reported stress levels, although the content was the same in the internet-based intervention. Interventions in both studies appear to be based on behavioral and developmental models relating to ADHD; however, it may be helpful for future research to establish whether videoconferencing is the most effective way to deliver training to carers (in comparison, eg, with a website, where carers are able to work through and revisit the materials at their own pace).

### *Mixed Mental Health Difficulties*

A total of 5 studies featured internet-based interventions for carers of individuals with mixed mental health difficulties, 3 of which [26,28,30] were RCTs, and 2 were prepost comparisons [27,29], with some evidence of reducing carer distress. These interventions were either codesigned with potential intervention users [29,30] or based pragmatically on use of mindfulness [26-28]—a technique that has been widely utilized across a range of population groups, but on the other hand, it is a technique that may not address the difficulties specific to the caring role; nonetheless, it was effective here. In future, it may be useful to compare disorder-specific carer interventions with more generic ones, potentially applicable to carers of people with a broad range of disorders to assess their relative merits in relation to their ability to improve carer outcomes.

## Stroke

A total of 2 of the 4 studies providing a internet-based intervention for carers of stroke survivors were moderate- and high-quality RCTs, respectively [32,33], with the remaining 2 studies [31,34] comprising relatively small, low-quality prepost comparison studies. We can only speculate on the reasons for this disparity in findings. The 2 studies [31,33] that found a positive impact of their interventions on carer mental health both delivered information to participants in a sequential, modular way. Such relatively structured intervention may be more accessible for older carers, who may have less experience in navigating websites. Of the studies that failed to find intervention effects on carer mental health, 1 [34] included carers of individuals with dementia, as well as carers of stroke survivors, which may have resulted in the content of the intervention being less tailored to the specific needs of stroke carers. Of note, both of the studies that did not show intervention effects in relation to carer mental health had other positive effects. In 1 of these, carers perceived greater social support following the internet-based intervention [34]. In the other study, stroke survivors whose carer received a internet-based intervention required fewer emergency department visits and fewer hospital readmissions than those whose carer did not, reducing the burden on the health service [32], which perhaps explains the lack of improvement in carer mental health. Further research in this area seems pressing, as in recent years, the number of stroke survivors has increased [66], with the vast majority of them living at home [67]. The informal carers of stroke survivors have been found to have high levels of depressive symptoms [68] and burden [69]. Thus, there is a considerable need for innovative interventions to support this population of carers. Future research in this area should seek to explore which particular aspect of a internet-based intervention leads to specific improvement across a wide variety of carer outcomes.

### Dementia

A total of 22 studies were found to investigate the effectiveness of internet-based interventions for carers of individuals with dementia (more than any other disorder discussed within this review), reflecting the size of this growing problem [70], the severity and range of behavioral and psychological symptoms that carers have to deal with in their loved one [71], and the corresponding severity of carer distress [14]. Findings from this review are largely in line with previous systematic reviews focused specifically on internet-based interventions for carers of people with dementia [72,73]. Although the results are mixed and this area of study would greatly benefit from more high-quality research, the evidence suggests that internet-based interventions may be useful in improving carer well-being and mental health.

There may be a number of reasons why these studies found mixed results regarding carer mental health; it may be the case that, over time, as the dementia sufferers' condition worsens, their carer experiences a greater sense of burden and related symptoms of mental distress. Alternatively, acquiring increased knowledge about the typical symptoms, course, and progressive nature of dementia through a internet-based intervention may negatively affect carers' mental health, especially if this is not buffered by having sufficient opportunity to have sensitive in-person discussions with a health professional. It is also of interest to note the disparate methods by which the interventions for dementia carers were developed. Although some appeared to be based to some extent on existing theories of, for example, stress and coping [35,74,75] or the transitions theory [39,50,76], the content of others was derived from interviews with carers and reviews of the relevant literature [48]. Future research in this area may want to investigate the feasibility of developing a model of carer distress specific to those supporting someone with dementia, on the basis of existing theories, in addition to being coproduced with carers. People with dementia are often cared for by their older spouses. For example, in a study of over 3800 dementia carers, the average age was found to be 63.3 years [77]. Thus, these carers may be less familiar with internet usage than younger people. As of 2011, only 41% of adults aged 65 years and above used the internet, in comparison to 94% of adults aged 18-29 years [78]. In addition, they may have visual or hearing impairments. Therefore, it is important to consider the appropriateness of internet-based interventions for this particular population and how such internet-based interventions may benefit from being altered to fit the specific needs of older carers. Several of the studies in this review used videoconferencing technology as part of their intervention—future research may wish to establish whether being able to see other carers or clinicians is more helpful to this population than interacting with a computer screen alone, as findings from this review are inconclusive with regard to this issue.

As noted above, there is some evidence that shorter interventions appeared to be more effective than those of longer duration in terms of their ability to reduce carer distress. The apparent effectiveness of shorter interventions in comparison to those of longer duration may have something to do with the nature of the illness. Over the course of the longer interventions, the symptoms of the person with dementia are likely to worsen, causing the person's carer to experience elevated levels of stress and burden.

### Traumatic Brain Injury

We identified 9 studies, all of which were RCTs of either moderate or high quality, evaluating a internet-based intervention for carers of people with TBI. A total of 8 of these 9 studies found some positive impact of the intervention on carer mental health. The majority of the TBI-carer interventions identified comprised family problem solving therapy. Thus, these interventions may have been more homogenous, compared with those developed for carers of people with other psychiatric or neurological disorders (where interventions targeting the same populations have been developed from a wider range of sources, models, and theories). Furthermore, 8 of the 9 studies followed a similar structure, comprising self-guided, modular interventions, accompanied by internet-based interaction with a therapist. In future, it would be useful to establish whether either or both of these aspects—a structured, modular program (as opposed to a site containing links, which the carer is required to navigate without guidance) or support by a clinician—are particularly associated with more positive psychological outcomes in carers. As most current studies on carer interventions for people with TBI have focused on children and adolescents, future research should seek to address the needs of carers of adults with TBI.

### General Discussion

The field of internet-based mental health interventions is still relatively new, and it continues to develop rapidly, including the recent recognition of the potential of internet-based interventions specifically aimed at carers. A previous review of "telehealth" (video, internet-based, telephone based, and telemetry or remote monitoring) interventions for family carers found that a majority of interventions were satisfactory to carers, and they were associated with significant improvements to carer outcomes [79], indicating the possible viability and effectiveness of technology-based support or training. This review specifically explored the effectiveness of internet-based interventions, and it discovered largely mixed findings with regard to the impact of interventions on carer mental health and well-being, with the exception of those aimed at carer of children and adolescents with TBI, almost all of which were found to have a positive impact on carer outcomes. Studies in this review focus on carers of individuals with a range of difficulties, including psychiatric disorders, neurological disorders, and brain injury, with diverse challenges for carers. Therefore, unsurprisingly, the format, content, and nature of interventions used here vary widely. Nonetheless, the evidence-based relating to some areas (dementia, TBI) is more extensive (with 6 and 5 large-scale RCTs, >100 participants, respectively) than in other areas (stroke, psychiatric disorders), where fewer large-scale RCTs have been conducted. Across almost all areas (with TBI a notable exception), findings in relation to reductions of carer distress are somewhat mixed. This may largely have to do with differences in populations and aspects of study design. In addition, not all of the studies presented focused primarily on carer outcomes, and several interventions with little or no impact on carer distress had other benefits (eg, improvements in patient outcomes).

Currently, there is not enough evidence to conclude whether interventions specifically designed with a particular disorder in mind, based on a clear model of carer distress, have advantages over more generic interventions. Of note, in the area of TBI, many of the successful interventions utilized a problem-solving approach, and in the area of mixed mental health problems, mindfulness approaches were successfully used.

A further aspect of internet-based interventions for which the evidence is also currently mixed is the impact of guidance or support. Several interventions included in this review include some guidance; however, this varied widely among studies, making it difficult to draw conclusions about whether guided interventions are superior to nonguided ones and which aspects of guidance (mode of delivery, training, or expertise of guides) are most important. A previous systematic review found guidance to be a beneficial aspect of internet-based interventions for mental health; however, methodological issues and the small number studies included in the review make it difficult to analyze these findings [80]. When considering the effectiveness of internet-based interventions, it is important to consider how an intervention delivered via the internet may differ from face-to-face treatment. Surprisingly, this review found only 2 studies that directly compared internet-based with face-to-face interventions. In a small study (n=37), carers of children with ADHD who received face-to-face training were found to have improvements in their levels of stress and strain, in comparison to those receiving the same content delivered on the Web [25]. Carers of individuals with mild dementia who received a internet-based intervention were found to have significant improvements in their quality of life in comparison to those who received care as usual (specified as "infrequent counseling"), but no difference was found between the 2 groups on the other 3 relevant outcome measures [40]. Multiple studies reported control groups that received "care as usual," but they did not specify what this comprised. Owing to the very small number of studies that did report a comparison between internet-based and face-to-face treatment, it is not possible to draw conclusions about how they compare in terms of impact on carer mental health or whether there are particular subgroups of carers who may benefit from one as opposed to the other. When attempting to explore the differences in response to the interventions discussed within this review, it may also be helpful to consider the possible applicability of Caregiver Identity Theory [7] to the responses of individual carers. Carer Identity Theory hypothesizes that the most significant factor influencing levels of distress is the disparity between carers' responsibilities and their perceived identity standard. If this is indeed the case, then it may be important to assess the individual carer, the carer's perceived burden, and the carer's changing relationship with the person for whom they care to determine what type and intensity of intervention may most benefit them. A further possible factor in explaining the differences observed in the efficacy of the different Web interventions presented here may be the wide range of initiatives encompassed by the phrase "internet-based" (eg, interactive websites accessed in the carers' own home vs videoconferencing chat, accessed from a local clinic). For some of the older studies included in this review, the technology utilized to provide the intervention may now be considered to be out of date. As the development of technology continues at an increasing pace, what may currently constitute a typical internet-based intervention may appear outdated in just a few years.

### Limitations

Several key limitations need to be noted. Although the majority of studies (29/46) were RCTs, a significant proportion comprised single-arm, prepost trials, with no randomization or control group comparisons. Quality varied widely across all studies, with some of the studies lacking key information on randomization methodology, blinding of assessors, and not having an accessible study protocol. This increases the difficulty of assessing the risk of bias within studies, and this makes it harder to interpret the effectiveness of the intervention reported. Overall, 16 RCTs and 0 non-RCTs were rated as high quality, 11 RCTs and 6 non-RCTs were rated as moderate quality, and 2 RCTs and 11 non-RCTs were rated as low quality, with study quality also varying across disorder type, making it hard to compare studies that were investigating the same type of disorder. Although we have reported all findings relating to carer mental health that were reported in the original study papers, a majority of the studies were not found to have a published paper detailing the study protocol, meaning that it was not always possible to confirm the absence of publication or selective reporting bias. Owing to the wide range of outcomes employed across studies to measure change in carer mental health and related constructs, we were unable to conduct any meta-analyses, making it more difficult to interpret and compare findings across studies. In addition, the diverse range of formats used to deliver the interventions and differing levels and types of guidance offered made it harder to compare the findings, and this may have had an impact on how engaging they are to carers.

### Future Considerations

Although the studies detailed above present a mixed picture regarding the overall effect of internet-based interventions on carer mental health, participating carers consistently reported that they found internet-based interventions to be highly acceptable and easy to utilize where these data were gathered. This may indicate that carers are willing to integrate internet-based interventions into their daily lives, and future research should attempt to establish the most effective way of delivering content to have the greatest impact possible on carer mental health. Of the 46 studies identified within this review, only 2 reported a direct comprising internet-based versus face-to-face interventions, of which only 1 study compared the same content delivered via the 2 different modalities. Future research should seek to establish whether the method of delivery of an intervention has an impact on carer outcomes and whether there are specific subgroups of carers who may benefit more from one than the other. As described above, internet-based interventions can comprise many different components, and it would be beneficial for future research to establish the degree to which particular elements contribute to forming an effective intervention. For example, it would be useful to be able to establish the active components of a given intervention; therefore, findings can be more easily accumulated and compared across interventions, improving the evidence base regarding which components of an intervention lead to desired behavior change [81]. The possibility of establishing specific models of carer distress on which interventions can be built should also be taken into consideration, in addition to consulting additional sources of information (such as focus groups, expert opinion, and reviews of the relevant literature). Future work into the impact of caring carer-interventions should consider taking a lifespan approach when considering the challenges of caring, both in terms of the carer life stage (eg, the differing requirements of young carers and older spousal carers) and that of the individual being cared for, as well as how these challenges can change and develop over time. Consideration (including with regard to cost effectiveness) should also be given to the possibility of blended care, where carers would receive some face-to-face clinician contact, in addition to accessing a internet-based intervention.

## Conclusions

The emerging field of internet-based interventions for carers of individuals with psychiatric disorders, neurological disorders, and brain injury offers exciting possibilities for providing support to a population that may otherwise find it difficult to access help, by giving them the option of accessing a range of relevant interventions in a more flexible way, or as part of a stepped-care program. Although findings from existing studies are mixed with regard to evidence of the efficacy of internet-based interventions, they show promise in terms of both effectiveness and acceptability, and further research into this area may establish the most effective ways in which internet-based interventions for carers can be utilized.

## Acknowledgments

The authors would like to thank the European Commission Horizon 2020 Programme and the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre (BRC) for providing the financial support that enabled this research to be conducted. LS is supported by a PhD studentship, funded by the European Commission-Horizon 2020 programme. RP is supported by an NIHR Maudsley BRC Studentship. US is supported by a Senior Investigator award from the NIHR and receives salary support from the NIHR BRC for Mental Health, South London and Maudsley National Health Service (NHS) Foundation Trust and Institute of Psychiatry, Psychology and Neuroscience, King's College London. The views expressed herein are not those of NIHR or the NHS. The funders played no role in the study design, data collection, interpretation and analysis, writing of the report, or the decision to submit the paper for publication. They accept no responsibility for the contents.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Risk of bias of included studies.

[PDF File (Adobe PDF File), 118KB - jmir_v21i7e10876_app1.pdf ]

## Multimedia Appendix 2

Summary of studies: carers of individuals with psychiatric disorders.

[PDF File (Adobe PDF File), 157KB - jmir_v21i7e10876_app2.pdf ]

## Multimedia Appendix 3

Summary of studies: carers of individuals who have survived a stroke.

[PDF File (Adobe PDF File), 110KB - jmir_v21i7e10876_app3.pdf ]

## Multimedia Appendix 4

Summary of studies: carers of individuals with dementia (studies with a control group).

[PDF File (Adobe PDF File), 159KB - jmir_v21i7e10876_app4.pdf ]

## Multimedia Appendix 5

Summary of studies: carers of individuals with dementia (studies without a control group).

[PDF File (Adobe PDF File), 140KB - jmir_v21i7e10876_app5.pdf ]

## Multimedia Appendix 6

Summary of studies: carers of individuals with traumatic brain injury.

[PDF File (Adobe PDF File), 138KB - jmir_v21i7e10876_app6.pdf ]

## References


1. Buckner L, Yeandle S. Carers UK. 2015. Valuing Carers 2015 – the rising value of carers' support URL: https://www.carersuk.org/for-professionals/policy/policy-library/valuing-carers-2015 [accessed 2019-06-10]
2. Brodaty H, Green A, Koschera A. Meta-analysis of psychosocial interventions for caregivers of people with dementia. J Am Geriatr Soc 2003 May;51(5):657-664. [Medline: 12752841]
3. Jungbauer J, Wittmund B, Dietrich S, Angermeyer MC. The disregarded caregivers: subjective burden in spouses of schizophrenia patients. Schizophr Bull 2004;30(3):665-675. [Medline: 15631259]
4. Martin Y, Gilbert P, McEwan K, Irons C. The relation of entrapment, shame and guilt to depression, in carers of people with dementia. Aging Ment Health 2006 Mar;10(2):101-106. [doi: 10.1080/13607860500311953] [Medline: 16517484]
5. Baumgarten M, Battista RN, Infante-Rivard C, Hanley JA, Becker R, Gauthier S. The psychological and physical health of family members caring for an elderly person with dementia. J Clin Epidemiol 1992 Jan;45(1):61-70. [Medline: 1738013]
6. Magliano L, Fiorillo A, De Rosa C, Malangone C, Maj M, National Mental Health Project Working Group. Family burden in long-term diseases: a comparative study in schizophrenia vs physical disorders. Soc Sci Med 2005 Jul;61(2):313-322. [doi: 10.1016/j.socscimed.2004.11.064] [Medline: 15893048]
7. Montgomery R, Kosloski K. Caregiving as a process of changing identity: implications for caregiver support. Generations 2009:47-52 https://www.ingentaconnect.com/content/asag/gen/2009/00000033/00000001/art00008.
8. Yesufu-Udechuku A, Harrison B, Mayo-Wilson E, Young N, Woodhams P, Shiers D, et al. Interventions to improve the experience of caring for people with severe mental illness: systematic review and meta-analysis. Br J Psychiatry 2015 Apr;206(4):268-274. [doi: 10.1192/bjp.bp.114.147561] [Medline: 25833867]
9. Blom MM, Bosmans JE, Cuijpers P, Zarit SH, Pot AM. Effectiveness and cost-effectiveness of an internet intervention for family caregivers of people with dementia: design of a randomized controlled trial. BMC Psychiatry 2013 Jan 10;13:17 [FREE Full text] [doi: 10.1186/1471-244X-13-17] [Medline: 23305463]
10. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
11. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs technician assistance. PLoS One 2010;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
12. Knaevelsrud C, Maercker A. Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial. BMC Psychiatry 2007;7:13 [FREE Full text] [doi: 10.1186/1471-244X-7-13] [Medline: 17442125]
13. Sánchez-Ortiz VC, Munro C, Stahl D, House J, Startup H, Treasure J, et al. A randomized controlled trial of internet-based cognitive-behavioural therapy for bulimia nervosa or related disorders in a student population. Psychol Med 2011 Feb;41(2):407-417. [doi: 10.1017/S0033291710000711] [Medline: 20406523]
14. Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003 Jun;18(2):250-267. [Medline: 12825775]
15. Hu C, Kung S, Rummans TA, Clark MM, Lapid MI. Reducing caregiver stress with internet-based interventions: a systematic review of open-label and randomized controlled trials. J Am Med Inform Assoc 2014 Aug 14:e194-e209. [doi: 10.1136/amiajnl-2014-002817] [Medline: 25125686]
16. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-based interventions to improve mental health, general caregiving outcomes, and general health for informal caregivers of adults with chronic conditions living in the community: rapid evidence review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]
17. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. J Med Internet Res 2015;17(5):e117 [FREE Full text] [doi: 10.2196/jmir.3717] [Medline: 25967983]
18. Reese RJ, Slone NC, Soares N, Sprang R. Telehealth for underserved families: an evidence-based parenting program. Psychol Serv 2012 Aug;9(3):320-322. [doi: 10.1037/a0026193] [Medline: 22867126]
19. Brockmeyer T, Friederich H, Schmidt U. Advances in the treatment of anorexia nervosa: a review of established and emerging interventions. Psychol Med 2018 Dec;48(8):1228-1256. [doi: 10.1017/S0033291717002604] [Medline: 28889819]
20. Grover M, Naumann U, Mohammad-Dar L, Glennon D, Ringwood S, Eisler I, et al. A randomized controlled trial of an Internet-based cognitive-behavioural skills package for carers of people with anorexia nervosa. Psychol Med 2011 Dec;41(12):2581-2591. [doi: 10.1017/S0033291711000766] [Medline: 21733215]

21. Hoyle D, Slater J, Williams C, Schmidt U, Wade TD. Evaluation of a web-based skills intervention for carers of people with anorexia nervosa: a randomized controlled trial. Int J Eat Disord 2013 Sep;46(6):634-638. [doi: 10.1002/eat.22144] [Medline: 23712500]
22. Bijker L, Kleiboer A, Riper H, Cuijpers P, Donker T. A pilot randomized controlled trial of E-care for caregivers: an internet intervention for caregivers of depressed patients. Internet Interv 2017 Sep;9:88-99 [FREE Full text] [doi: 10.1016/j.invent.2017.06.003] [Medline: 30135842]
23. Rotondi AJ, Haas GL, Anderson CM, Newhill CE, Spring MB, Ganguli R, et al. A clinical trial to test the feasibility of a telehealth psychoeducational intervention for persons with schizophrenia and their families: intervention and 3-month findings. Rehabil Psychol 2005 Nov;50(4):325-336 [FREE Full text] [doi: 10.1037/0090-5550.50.4.325] [Medline: 26321774]
24. Glynn SM, Randolph ET, Garrick T, Lui A. A proof of concept trial of an online psychoeducational program for relatives of both veterans and civilians living with schizophrenia. Psychiatr Rehabil J 2010;33(4):278-287. [doi: 10.2975/33.4.2010.278.287] [Medline: 20374986]
25. Tse YJ, McCarty CA, Stoep AV, Myers KM. Teletherapy delivery of caregiver behavior training for children with attention-deficit hyperactivity disorder. Telemed J E Health 2015 Jun;21(6):451-458 [FREE Full text] [doi: 10.1089/tmj.2014.0132] [Medline: 25719609]
26. Stjernswärd S, Hansson L. Effectiveness and usability of a web-based mindfulness intervention for families living with mental illness. Mindfulness (N Y) 2017;8(3):751-764 [FREE Full text] [doi: 10.1007/s12671-016-0653-2] [Medline: 28515801]
27. Stjernswärd S, Hansson L. Outcome of a web-based mindfulness intervention for families living with mental illness - a feasibility study. Inform Health Soc Care 2017 Jan;42(1):97-108. [doi: 10.1080/17538157.2016.1177533] [Medline: 27245198]
28. Stjernswärd S, Hansson L. Effectiveness and usability of a web-based mindfulness intervention for caregivers of people with mental or somatic illness. A randomized controlled trial. Internet Interv 2018 Jun;12:46-56 [FREE Full text] [doi: 10.1016/j.invent.2018.03.004] [Medline: 30135768]
29. Gleeson J, Lederman R, Koval P, Wadley G, Bendall S, Cotton S, et al. Moderated online social therapy: a model for reducing stress in carers of young people diagnosed with mental health disorders. Front Psychol 2017;8:485 [FREE Full text] [doi: 10.3389/fpsyg.2017.00485] [Medline: 28421012]
30. Ali L, Krevers B, Sjöström N, Skärsäter I. Effectiveness of web-based versus folder support interventions for young informal carers of persons with mental illness: a randomized controlled trial. Patient Educ Couns 2014 Mar;94(3):362-371. [doi: 10.1016/j.pec.2013.10.020] [Medline: 24341963]
31. Graf R, LeLaurin J, Schmitzberger M, Freytes IM, Orozco T, Dang S, et al. The stroke caregiving trajectory in relation to caregiver depressive symptoms, burden, and intervention outcomes. Top Stroke Rehabil 2017 Oct;24(7):488-495. [doi: 10.1080/10749357.2017.1338371] [Medline: 28618848]
32. Pierce LL, Steiner VL, Khuder SA, Govoni AL, Horn LJ. The effect of a web-based stroke intervention on carers' well-being and survivors' use of healthcare services. Disabil Rehabil 2009;31(20):1676-1684. [doi: 10.1080/09638280902751972] [Medline: 19479528]
33. Smith GC, Egbert N, Dellman-Jenkins M, Nanna K, Palmieri PA. Reducing depression in stroke survivors and their informal caregivers: a randomized clinical trial of a web-based intervention. Rehabil Psychol 2012 Aug;57(3):196-206 [FREE Full text] [doi: 10.1037/a0029587] [Medline: 22946607]
34. 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]
35. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. Gerontologist 2005 Dec;45(6):793-801. [Medline: 16326661]
36. Eisdorfer C, Czaja SJ, Loewenstein DA, Rubert MP, Argüelles S, Mitrani VB, et al. The effect of a family therapy and technology-based intervention on caregiver depression. Gerontologist 2003 Aug;43(4):521-531 [FREE Full text] [Medline: 12937331]
37. van Mierlo LD, Meiland FJ, van de Ven PM, van Hout HP, Dröes RM. Evaluation of DEM-DISC, customized e-advice on health and social support services for informal carers and case managers of people with dementia; a cluster randomized trial. Int Psychogeriatr 2015 Aug;27(8):1365-1378. [doi: 10.1017/S1041610215000423] [Medline: 25872457]
38. Pagán-Ortiz ME, Cortés DE, Rudloff N, Weitzman P, Levkoff S. Use of an online community to provide support to caregivers of people with dementia. J Gerontol Soc Work 2014;57(6-7):694-709 [FREE Full text] [doi: 10.1080/01634372.2014.901998] [Medline: 24689359]
39. Duggleby W, Ploeg J, McAiney C, Peacock S, Fisher K, Ghosh S, et al. Web-based intervention for family carers of persons with dementia and multiple chronic conditions (My Tools 4 Care): pragmatic randomized controlled trial. J Med Internet Res 2018 Dec 29;20(6):e10484 [FREE Full text] [doi: 10.2196/10484] [Medline: 29959111]
40. Boots LM, de Vugt ME, Kempen GI, Verhey FR. Effectiveness of a blended care self-management program for caregivers of people with early-stage dementia (partner in balance): randomized controlled trial. J Med Internet Res 2018 Jul 13;20(7):e10017 [FREE Full text] [doi: 10.2196/10017] [Medline: 30006327]

41. Blom MM, Zarit SH, Zwaaftink RB, Cuijpers P, Pot AM. Effectiveness of an Internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. PLoS One 2015;10(2):e0116622 [FREE Full text] [doi: 10.1371/journal.pone.0116622] [Medline: 25679228]
42. Czaja SJ, Loewenstein D, Schulz R, Nair SN, Perdomo D. A videophone psychosocial intervention for dementia caregivers. Am J Geriatr Psychiatry 2013 Nov;21(11):1071-1081. [doi: 10.1016/j.jagp.2013.02.019] [Medline: 23831174]
43. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian Tzuang Y, et al. Exploring the effectiveness of an internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. Aging Ment Health 2013;17(5):544-554 [FREE Full text] [doi: 10.1080/13607863.2013.775641] [Medline: 23461355]
44. Marziali E, Donahue P. Caring for others: internet video-conferencing group intervention for family caregivers of older adults with neurodegenerative disease. Gerontologist 2006 Jun;46(3):398-403 [FREE Full text] [Medline: 16731880]
45. Torkamani M, McDonald L, Aguayo IS, Kanios C, Katsanou MN, Madeley L, ALADDIN Collaborative Group. A randomized controlled pilot study to evaluate a technology platform for the assisted living of people with dementia and their carers. J Alzheimers Dis 2014;41(2):515-523. [doi: 10.3233/JAD-132156] [Medline: 24643137]
46. Griffiths PC, Whitney MK, Kovaleva M, Hepburn K. Development and implementation of tele-savvy for dementia caregivers: a department of veterans affairs clinical demonstration project. Gerontologist 2016 Feb;56(1):145-154. [doi: 10.1093/geront/gnv123] [Medline: 26566806]
47. Kwok T, Au A, Wong B, Ip I, Mak V, Ho F. Effectiveness of online cognitive behavioral therapy on family caregivers of people with dementia. Clin Interv Aging 2014;9:631-636 [FREE Full text] [doi: 10.2147/CIA.S56337] [Medline: 24748781]
48. Lorig K, Thompson-Gallagher D, Traylor L, Ritter PL, Laurent DD, Plant K, et al. Building better caregivers: a pilot online support workshop for family caregivers of cognitively impaired adults. J Appl Gerontol 2010 Dec 15;31(3):423-437. [doi: 10.1177/0733464810389806]
49. Bateman DR, Brady E, Wilkerson D, Yi E, Karanam Y, Callahan CM. Comparing crowdsourcing and friendsourcing: a social media-based feasibility study to support Alzheimer disease caregivers. JMIR Res Protoc 2017 Apr 10;6(4):e56 [FREE Full text] [doi: 10.2196/resprot.6904] [Medline: 28396304]
50. Duggleby W, Jovel Ruiz K, Ploeg J, McAiney C, Peacock S, Nekolaichuk C, et al. Mixed-methods single-arm repeated measures study evaluating the feasibility of a web-based intervention to support family carers of persons with dementia in long-term care facilities. Pilot Feasibility Stud 2018;4:165 [FREE Full text] [doi: 10.1186/s40814-018-0356-7] [Medline: 30410783]
51. Chiu T, Marziali E, Colantonio A, Carswell A, Gruneir M, Tang M, et al. Internet-based caregiver support for Chinese Canadians taking care of a family member with alzheimer disease and related dementia. Can J Aging 2009 Dec;28(4):323-336. [doi: 10.1017/S0714980809990158] [Medline: 19925698]
52. Glueckauf RL, Ketterson TU, Loomis JS, Dages P. Online support and education for dementia caregivers: overview, utilization, and initial program evaluation. Telemed J E Health 2004;10(2):223-232. [Medline: 15319052]
53. Lindauer A, Croff R, Mincks K, Mattek N, Shofner SJ, Bouranis N, et al. "It Took the Stress out of Getting Help": The STAR-C-Telemedicine Mixed Methods Pilot. Care Wkly 2018;2:7-14 [FREE Full text] [doi: 10.14283/cw.2018.4] [Medline: 30393783]
54. Marziali E, Garcia LJ. Dementia caregivers' responses to 2 Internet-based intervention programs. Am J Alzheimers Dis Other Demen 2011 Feb;26(1):36-43. [doi: 10.1177/1533317510387586] [Medline: 21282276]
55. McLaughlin KA, Glang A, Beaver SV, Gau JM, Keen S. Web-based training in family advocacy. J Head Trauma Rehabil 2013;28(5):341-348 [FREE Full text] [doi: 10.1097/HTR.0b013e31824e1d43] [Medline: 22688211]
56. Petranovich CL, Wade SL, Taylor HG, Cassedy A, Stancin T, Kirkwood MW, et al. Long-term caregiver mental health outcomes following a predominately online intervention for adolescents with complicated mild to severe traumatic brain injury. J Pediatr Psychol 2015 Aug;40(7):680-688 [FREE Full text] [doi: 10.1093/jpepsy/jsv001] [Medline: 25682211]
57. Raj SP, Antonini TN, Oberjohn KS, Cassedy A, Makoroff KL, Wade SL. Web-based parenting skills program for pediatric traumatic brain injury reduces psychological distress among lower-income parents. J Head Trauma Rehabil 2015;30(5):347-356. [doi: 10.1097/HTR.0000000000000052] [Medline: 24842588]
58. Raj SP, Shultz EL, Zang H, Zhang N, Kirkwood MW, Taylor HG, et al. Effects of web-based parent training on caregiver functioning following pediatric traumatic brain injury: a randomized control trial. J Head Trauma Rehabil 2018;33(6):E19-E29. [doi: 10.1097/HTR.0000000000000388] [Medline: 29601344]
59. Wade SL, Walz NC, Carey J, McMullen KM, Cass J, Mark E, et al. A randomized trial of teen online problem solving: efficacy in improving caregiver outcomes after brain injury. Health Psychol 2012 Nov;31(6):767-776. [doi: 10.1037/a0028440] [Medline: 22746261]
60. Wade SL, Karver CL, Taylor HG, Cassedy A, Stancin T, Kirkwood MW, et al. Counselor-assisted problem solving improves caregiver efficacy following adolescent brain injury. Rehabil Psychol 2014 Feb;59(1):1-9 [FREE Full text] [doi: 10.1037/a0034911] [Medline: 24611923]
61. Carey JC, Wade SL, Wolfe CR. Lessons learned: the effect of prior technology use on web-based interventions. Cyberpsychol Behav 2008 Apr;11(2):188-195 [FREE Full text] [doi: 10.1089/cpb.2007.0025] [Medline: 18422412]
62. Wade SL, Carey J, Wolfe CR. An online family intervention to reduce parental distress following pediatric brain injury. J Consult Clin Psychol 2006 Jun;74(3):445-454. [doi: 10.1037/0022-006X.74.3.445] [Medline: 16822102]

63. Wade SL, Walz NC, Carey JC, Williams KM. Preliminary efficacy of a web-based family problem-solving treatment program for adolescents with traumatic brain injury. J Head Trauma Rehabil 2008;23(6):369-377. [doi: 10.1097/01.HTR.0000341432.67251.48] [Medline: 19033829]
64. Treasure J, Whitaker W, Whitney J, Schmidt U. Working with families of adults with anorexia nervosa. J Family Therapy 2005 May;27(2):158-170. [doi: 10.1111/j.1467-6427.2005.00308.x]
65. Weisz JR, Chorpita BF, Palinkas LA, Schoenwald SK, Miranda J, Bearman SK, Research Network on Youth Mental Health. Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: a randomized effectiveness trial. Arch Gen Psychiatry 2012 Mar;69(3):274-282. [doi: 10.1001/archgenpsychiatry.2011.147] [Medline: 22065252]
66. Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, Global Burden of Diseases, Injuries, Risk Factors Study 2010 (GBD 2010) the GBD Stroke Experts Group. Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. Lancet 2014 Jan 18;383(9913):245-254 [FREE Full text] [Medline: 24449944]
67. Crichton SL, Bray BD, McKevitt C, Rudd AG, Wolfe CD. Patient outcomes up to 15 years after stroke: survival, disability, quality of life, cognition and mental health. J Neurol Neurosurg Psychiatry 2016 Dec;87(10):1091-1098. [doi: 10.1136/jnnp-2016-313361] [Medline: 27451353]
68. Visser-Meily A, Post M, van de Port I, Maas C, Forstberg-Wärleby G, Lindeman E. Psychosocial functioning of spouses of patients with stroke from initial inpatient rehabilitation to 3 years poststroke: course and relations with coping strategies. Stroke 2009 Apr;40(4):1399-1404 [FREE Full text] [doi: 10.1161/STROKEAHA.108.516682] [Medline: 19095973]
69. Jaracz K, Grabowska-Fudala B, Górna K, Jaracz J, Moczko J, Kozubski W. Burden in caregivers of long-term stroke survivors: Prevalence and determinants at 6 months and 5 years after stroke. Patient Educ Couns 2015 Aug;98(8):1011-1016. [doi: 10.1016/j.pec.2015.04.008] [Medline: 25952926]
70. Prince M. Alzheimer's Disease International. 2015. World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends URL: https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf [accessed 2019-06-10]
71. Savva GM, Zaccai J, Matthews FE, Davidson JE, McKeith I, Brayne C. Prevalence, correlates and course of behavioural and psychological symptoms of dementia in the population. Br J Psychiatry 2009 Mar;194(3):212-219. [doi: 10.1192/bjp.bp.108.049619] [Medline: 19252147]
72. Boots LM, de Vugt ME, van Knippenberg RJ, Kempen GI, Verhey FR. A systematic review of internet-based supportive interventions for caregivers of patients with dementia. Int J Geriatr Psychiatry 2014 Apr;29(4):331-344. [doi: 10.1002/gps.4016] [Medline: 23963684]
73. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review. J Med Internet Res 2018 Dec 12;20(6):e216 [FREE Full text] [doi: 10.2196/jmir.9548] [Medline: 29895512]
74. Lazarus RS, Folkman S. Stress, Appraisal, And Coping. New York: Springer Publishing Company; 1984.
75. Gallagher-Thompson D, Solano N, McGee JS, Krisztal E, Kaye J, Coon DW, et al. Coping With Caregiving: Reducing Stress and Improving Your Quality of Life. Stanford University: Stanford University School of Medicine and VA Palo Alto Health Care System; 2002.
76. Meleis A. Transitions Theory. In: Nursing Theories And Nursing Practice. Philadelphia: FA Davis Company; 2015.
77. Yaffe K, Fox P, Newcomer R, Sands L, Lindquist K, Dane K, et al. Patient and caregiver characteristics and nursing home placement in patients with dementia. J Am Med Assoc 2002 Apr 24;287(16):2090-2097. [Medline: 11966383]
78. Zickuhr K, Smith A. Pew Internet. 2012. Digital differences URL: https://www.pewinternet.org/2012/04/13/digital-differences/ [accessed 2019-06-10]
79. Chi N, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2015 Jan;21(1):37-44 [FREE Full text] [doi: 10.1177/1357633X14562734] [Medline: 25475220]
80. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
81. 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]


## Abbreviations

**ADHD:** attention-deficit hyperactivity disorder
**AN:** anorexia nervosa
**BRC:** Biomedical Research Centre
**CBT:** cognitive behavioral therapy
**NHS:** National Health Service
**NIHR:** National Institute for Health Research

**OAO:** Overcoming Anorexia Online
**RCT:** randomized controlled trial
**TBI:** traumatic brain injury

*Edited by G Eysenbach; submitted 26.04.18; peer-reviewed by J Jarvis, G Baumblatt; comments to author 30.08.18; revised version received 15.03.19; accepted 24.03.19; published 09.07.19.*

*Please cite as:*
*Spencer L, Potterton R, Allen K, Musiat P, Schmidt U*
*Internet-Based Interventions for Carers of Individuals With Psychiatric Disorders, Neurological Disorders, or Brain Injuries: Systematic Review*
*J Med Internet Res 2019;21(7):e10876*
*URL: https://www.jmir.org/2019/7/e10876/*
*doi:10.2196/10876*
*PMID:31290399*

Original Paper

# The Cost-Effectiveness of an Internet Intervention to Facilitate Mental Health Help-Seeking by Young Adults: Randomized Controlled Trial

Long Khanh-Dao Le[1], BPharm, MPH, MHHSM, PhD; Lena Sanci[2], MBBS, PhD; Mary Lou Chatterton[1], PharmD; Sylvia Kauer[2], PhD; Kerrie Buhagiar[3], PhD; Cathrine Mihalopoulos[1], PhD

[1]Deakin Health Economics, Institute for Health Transformation, School of Health and Social Development, Deakin University, Burwood, Australia
[2]Department of General Practice, University of Melbourne, Melbourne, Australia
[3]ReachOut Australia, NSW, Australia

**Corresponding Author:**
Long Khanh-Dao Le, BPharm, MPH, MHHSM, PhD
Deakin Health Economics, Institute for Health Transformation
School of Health and Social Development
Deakin University
221 Burwood Highway
Burwood, 3125
Australia
Phone: 61 392468383
Email: long.le@deakin.edu.au

## Abstract

**Background:** Little empirical evidence is available to support the effectiveness and cost-effectiveness of internet interventions to increase help-seeking behavior for mental health in young adults.

**Objective:** The aim of this study was to evaluate the cost-effectiveness of a Web-based mental health help-seeking navigation tool (*Link*) in comparison with usual help-seeking strategies.

**Methods:** A cost-utility analysis alongside the main randomized trial of *Link* was conducted from the Australian health care sector perspective. Young adults aged 18 to 25 years were randomized to the *Link* intervention (n=205) or usual care (n=208) with 1- and 3-month follow-ups. The primary outcome of this study was quality-adjusted life years (QALYs) measured by the assessment of quality of life–4D. Costs were calculated based on the self-reported resource use questionnaire and were reported in 2015 Australian dollars. Primary analyses were conducted as intention-to-treat and reported as incremental cost-effectiveness ratios. Completer analyses were conducted in a sensitivity analysis.

**Results:** Significantly more QALYs were gained in the intervention group than the control group (0.15 vs 0.14; $P$<.001). The intervention was associated with significantly lower health professional consultation costs at 1-month follow-up (mean costs Aus $98 vs Aus $162; $P$<.05). Costs of hospital services were lower at 3 months in the intervention arm (mean costs Aus $47 vs Aus $101); however, there was insufficient sample size to detect a significant difference between the groups. There were no statistically significant differences in the total costs between the 2 arms. Relative to the control group, those who received the intervention experienced 0.01 more QALYs (0.00-0.02) and had lower total health sector costs of Aus −$81 (Aus −$348 to Aus $186) over 3 months. The intervention was found to be more effective and less costly compared with usual help-seeking strategies. The intervention was 100% likely to be cost-effective below a willingness-to-pay value-for-money threshold of Aus $28,033 per QALY. Results were robust in the sensitivity analysis.

**Conclusions:** Our study found that the online youth mental health help-seeking Web service is a cost-effective intervention for young people aged 18 to 25 years compared with usual search strategies. Further research is required to confirm these results.

**Trial Registration:** Australian New Zealand Clinical Trials Registry ACTRN12614001223628; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=366731

***(J Med Internet Res 2019;21(7):e13065)*** doi:10.2196/13065

**KEYWORDS**

economic evaluation; cost effectiveness; mental health; help-seeking; internet intervention

# Introduction

## Background

Mental and substance use disorders are a leading cause of disability in children and young adults worldwide [1], making these diagnoses a significant public health concern. Mental disorders were also associated with substantial economic burden with an estimated total cost of Aus $12.7 billion annually within the Australian context [2]. Despite the significant effect of these conditions in young people, which may continue into adulthood, only 23.3% of young adults (aged 16-24 years) with a 12-month diagnosis of a mental disorder in Australia sought professional treatment for mental health problems [3].

Barriers to help-seeking and treatment for young people include stigma [4-7], embarrassment [5], poor mental health literacy [5,7], lack of knowledge about appropriate mental health services [6-8], and a preference for self-reliance [5,6] in addition to geographic barriers for those living in rural settings with limited access to resources [9,10]. E-mental health interventions delivered through internet or mobile phone technology show promise [11]; however, little empirical evidence is available to support the effectiveness and cost-effectiveness of these interventions to increase help-seeking behavior [12].

## Objective

To address these concerns, a randomized controlled trial (RCT) was conducted to evaluate the effectiveness and cost-effectiveness of a brief, internet-based, mental health help-seeking intervention, called *Link* compared with usual help-seeking strategies for young adults. The current analysis sought to answer whether an online help-seeking intervention for young adults was cost-effective compared with usual search practices from a health care sector perspective (defined as health care government expenditure plus health care out-of-pocket expenditure) within a 3-month follow-up.

# Methods

## Approval and Ethical Considerations

Ethics approval was obtained from the University of Melbourne Human Research Ethics Committee, reference #1341063.4, and Deakin University Human Research Ethics Committee, reference #2015-320. All participants consented to take part in this study via an online consent form.

## Study Design and Participants

This economic evaluation was conducted alongside the RCT. The study adheres to the Consolidated Health Economic Evaluation Reporting Standards Statement (CHEERS) checklist [13] (Multimedia Appendix 1).

The study was conducted entirely online. Participants were recruited by electronic direct mail, social media, online advertising, and snowballing, where participants were asked to share the link on the Facebook page with friends and family. Interested participants were directed from a link in the advertisements to the study website where they were provided with more information and a consenting procedure if meeting the eligibility criteria of being aged between 18 and 25 years and residing in Australia. Eligible participants provided informed consent by acknowledging that they had read the information statement by clicking a box, then clicking a separate box to indicate that they consented to participate in the Link Research Project. They then registered for the trial using their email address and a self-generated password. Immediately following registration, all participants completed the baseline survey sent through email including demographic information and the Kessler-10 (K10) measure of psychological distress. Participants were then stratified by responses on gender (male or female) and severity of psychological distress (K10>20), then randomized into parallel groups consisting of the intervention group (*Link*) or control group (usual search strategies) using a random allocation sequence generated internally by the QuON computer software [14]. Randomization was stratified by gender (male and female) and psychological distress (K10 score<20 and K10 score≥20) using random sequences of block sizes of 4, 6, or 8 within each stratum and an allocation ratio of 1:1. Online surveys were completed by all participants at baseline, postintervention, and 3-month follow-up. Survey measures included the positive affect and negative affect scale, barriers to adolescent help-seeking, stages of change questionnaire, K10, general help-seeking questionnaire, assessment of quality of life (AQoL)–4D, client satisfaction questionnaire, and the health service use questionnaire. Researchers and statisticians involved in the data analysis were blind to the allocation of participants until after data analysis was completed. Further information related to the trial can be found in the paper reporting the primary trial outcomes [15].

## Intervention Descriptions

### Intervention Arm

The *Link* intervention is an online Web-based mental health help-seeking tool designed to guide young adults to appropriate online and offline sources of mental health information and care. The *Link* design is underpinned by the theory of planned behavior [16] and the Help-Seeking Model [17]. The functionality of *Link* operationalizes the elements of these theories (attitudes toward help-seeking, subjective norms, perceived control of help-seeking, and intentions to seek help) toward encouraging help-seeking behavior [18]. In brief, *Link* has a 4-step process where (1) users select symptoms they experience, (2) rate how much they are affected by them, (3) choose their preferred way to receive help (face-to-face, online information, telephone, and online chat), and then (4) finally, click on service options presented by the program for more information on how to seek help within that service, including expected costs and website links or online directories. The feasibility of *Link* was trialed previously and found to be acceptable to young people [19].

### *Control Arm: Usual Search Strategies*

The control condition instructed the young adult participants to use their typical strategies to seek help both online and offline such as using internet search engines and face-to-face or phone services.

## Outcome Measures

### *Health-Related Quality of Life*

The AQoL-4D was used to measure health-related quality of life [20]. Originally developed as a generic multiattribute utility instrument designed for the evaluation of public health interventions including mental health [20], it originally consisted of 15 items spread out into 5 dimensions measuring illness, independent living, social relationships, physical senses, and psychological well-being. However, the illness subscale was not used in the scoring [20]. The AQoL-4D scoring algorithm, based on the multiattribute utility theory, weighs the items and then applies a multiplicative model to obtain an index, which is transformed into a utility scale [20]. Quality-adjusted life years (QALYs) were calculated over the time horizon of the study using the area under the curve method [21].

### *Costs*

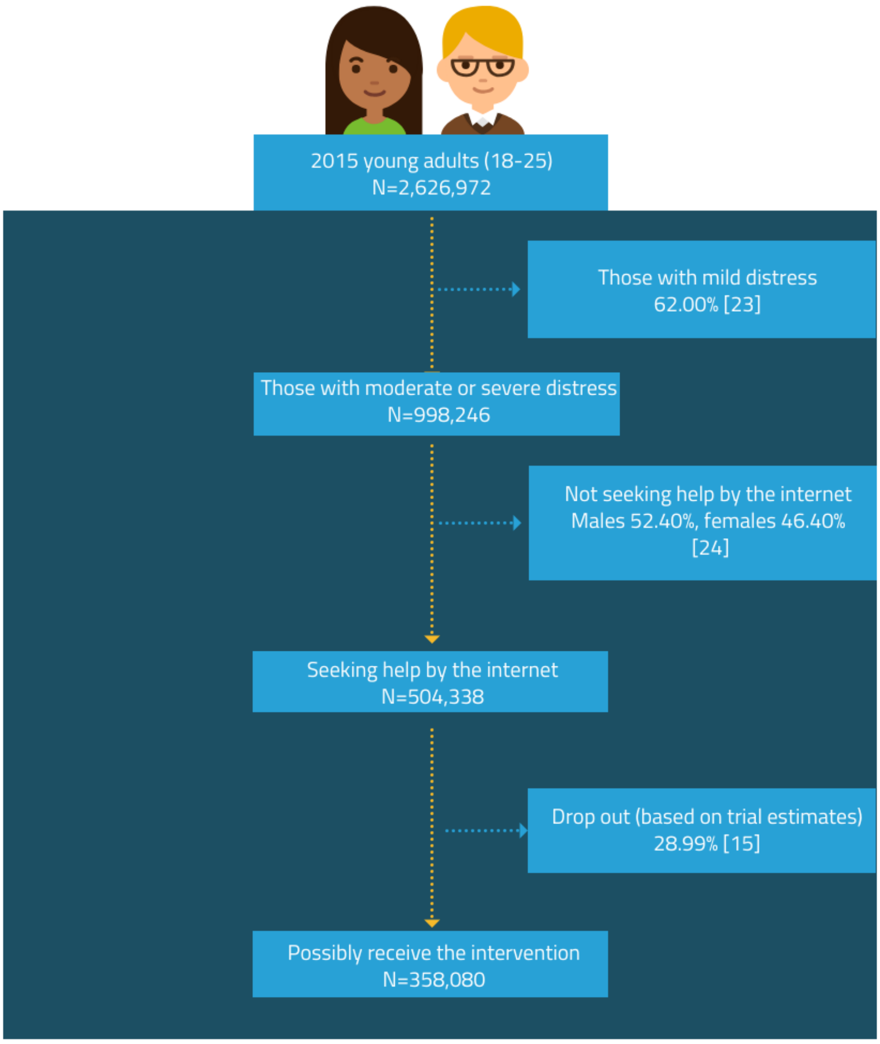
This economic evaluation adopted a health sector perspective, which included health care costs paid by the government and out-of-pocket costs paid by patients. All costs were expressed as 2015 Australian Dollars. No discount rate was applied because the time horizon of the study was 3 months.

#### Intervention Costs

Intervention costs comprised the intervention development costs and maintenance costs. Development costs were estimated from the details provided by the research team and included the planning, development, and production stages of the *Link* platform. The total projected cost for *Link* was Aus $1.74 million. The maintenance cost of the *Link* intervention included the time cost of 2 information technology staff (1 senior and 1 junior staff), in addition to the time cost of staff to update content and equipment costs. The total maintenance cost for *Link* was Aus $29,803 per year (or equivalent to Aus $2484 per month). To not overestimate the per-person costs (by assigning them only to trial participants), we estimated the number of people who are likely to receive the intervention when implemented within the Australian population using assumptions based on the published literature. The intervention pathway starts with young adults aged 18 to 25 years in the 2015 Australian population [22]. Despite no restriction of the intervention for young adults, we conservatively assumed that those with moderate or high mental health distress (measured by K10) are likely to have an interest in help-seeking for mental health problems [23]. For those people, approximately half were assumed to seek help through the internet based on the Mission Australia Youth Survey [24]. Furthermore, we also assumed a 29% dropout based on the dropout rate of this trial [15]. As a result, approximately 14% of Australian young adults were assumed to use the *Link* intervention (Figure 1).

This resulted in the average development cost per person for the *Link* program estimated at Aus $5.59, and the total average maintenance cost was estimated at Aus $0.04 per person per month. Therefore, the total intervention costs per person for the 3-month follow-up were estimated at approximately Aus $5.84.

**Figure 1.** Estimation of population eligibility for the Link intervention.




### Health Care Utilization Costs

Health care utilization was self-reported by participants at 1 and 3 months, retrospectively, using a resource use questionnaire (RUQ). The RUQ comprised questions on relevant health care services (eg, general practitioner [GP], psychologist, and/or mental health specialists or health experts), including the frequency of visits, payment methods (ie, out - of - pocket payments), outpatient care services (ie, nonadmitted hospital-based services), inpatient admissions, and medications. The different versions of the RUQ have been used in other trials in mental health [25]. The costs were calculated by multiplying the reported number of contacts by standard Australian unit costs. Unit costs for consultations (ie, GP, psychologist, psychiatrist, and allied health professionals) were sourced from the 2014 Medicare Benefit Schedule Book [22] and presented in Multimedia Appendix 2. Unit costs for medications adopted a weighted average of all available products containing the relevant active ingredient sourced from 2014 Pharmaceutical Benefit Schedule reports [26]. Hospital stays were costed using public sector average cost per separation through the Independent Hospital Pricing Authority, based on Australian Refined Diagnostic Related Group (AR-DRG) [27]. The specific AR-DRGs (for mental health symptoms) were chosen based on the self-reported reason and duration of stay.

The out-of-pocket costs reported in the RUQ for each service were considered in the health sector perspective. If the reported amount for a community-based health contact was outside of a plausible range, the maximum of out-of-pocket cost of Aus $447 was used based on the recommendation of the Australian Psychological Society [28]. For those who did not report out-of-pocket costs, we assumed that no out-of-pocket costs were incurred.

### Statistical Analysis

The primary analysis was performed using an intention-to-treat approach. All participants who were randomized were included in the analysis, and missing data were handled by multiple imputation by chained equations using predictive mean matching. The data were assumed to be missing at random by testing through a series of logistic regression analyses comparing participants' characteristics for those with and without missing endpoint data. At 1- and 3-month follow-ups, approximately 30% of participants had dropped out or did not complete the survey (29% in the intervention group vs 31% in control group). However, the maximum percentage of missing QALY and cost data was 40%. Thus, to ensure efficient and reproducible estimates, a total of 40 imputations were completed [29,30]. The estimates obtained from each imputed dataset were combined using Rubin's rules to generate an overall mean estimate of QALYs and costs. Rubin's rules ensure that the standard error reflects the variability within and across imputations.

General linear models (GLMs) were used to evaluate differences between group on total QALYs and total health sector costs. For the GLMs, a modified Park test was used to identify the appropriate *family*, whereas Pregibon link test, Pearson correlation test, and modified Hosmer-Lemeshow test were adopted to identify the *link function* [21]. GLM with log link and Gaussian family was conducted for QALYs. Given the large proportion of zero costs, 2-part models were used to evaluate the difference in components of the total costs including consultations, hospital, and medication costs between intervention and control groups as recommended in the literature [21]. We first modeled the probability that a person has any health care expenditures with a logit model using the full sample. Then we estimated a GLM on the subset of people who have any expenditures. The 2-part model allows for separate investigation of the effect of covariates on the extensive margin (logit model, if any expenditures) and on the intensive margin (GLM, amount of expenditures if any) [31,32]. GLM using log link and gamma family was used for cost variables as recommended by the International Society for Pharmacoeconomics and Outcome Research guidelines [33]. All regression analyses were adjusted by the utility scores at baseline, gender (male and female), baseline K10 scores, and the use of online searches for mental health services in the 2 weeks before study entry. The incremental difference in costs and QALYs between groups was estimated based on the 3-month data using seemingly unrelated regression model, combining estimates of mean coefficients and the covariance matrix as per Rubin's rules [34]. The regression coefficient on the treatment variable in the cost and QALY equations represents the incremental differences in costs and QALYs, respectively. The incremental cost-effectiveness ratio (ICER) was calculated as the ratio of these coefficients.

The bias-corrected CIs around the ICER were reported based on 3000 bootstrap simulations. The bootstrapped data were also plotted on a cost-effectiveness plane [35]. The threshold willingness-to-pay of Aus $28,033 per QALY gained was used to determine cost-effectiveness because this reflects the opportunity costs of decisions to publicly fund new health technologies in Australia [36]. In addition, a cost-effectiveness acceptability curve was constructed by calculating the probability of the intervention being cost-effective at different values of willingness-to-pay [37]. The probability of cost-effectiveness was estimated from combining mean coefficients and the covariance matrix from the seemingly unrelated regression model. The validity of this approach relies on the multivariate normality of the group-specific mean costs and QALYs [34]. This is appropriate with a sufficient sample size even when individual costs and QALYs are skewed [34,38].

Sensitivity analyses included a complete case analysis in which only participants who completed 1- and 3-month follow-ups were included. In addition, the development costs were varied to reflect different proportions of the population receiving the intervention if it was implemented in Australia. In particular, the proportion of people who would receive the intervention was varied from 2% to 17% of the Australian population.

All analyses were undertaken using Stata SE version 15.

## Results

### Overview

A total of 413 participants were randomized, with 205 allocated to *Link* and 208 allocated to the control group. Additional details regarding the study flow and Consort diagram are reported elsewhere [15]. The overall attrition rates were similar between the 2 study groups (71% *Link* vs 69% control group). Baseline characteristics were similar between the groups (Table 1), except a significantly greater proportion of participants in the intervention group carried out an online search of mental health services in the 2 weeks before randomization compared with the control group (38.5% vs 26%, $P$<.01).

**Table 1.** Baseline characteristics of the study population.

| Characteristics | *Link* intervention (n=205) | Control (n=208) |
|---|---|---|
| **Gender, n (%)** | | |
| Female | 171 (83.4) | 173 (83.2) |
| Other[a] | 3 (1.5) | 4 (1.9) |
| **Education, n (%)** | | |
| Completed secondary school | 104 (50.7) | 99 (47.6) |
| Higher education | 90 (43.9) | 95 (45.7) |
| **Working status, n (%)[b]** | | |
| Yes | 107 (52.2) | 117 (56.3) |
| **Absent study days, n (%)** | | |
| Yes | 58 (28.3) | 71 (34.1) |
| **K10[c] categories, n (%)** | | |
| Mild | 28 (13.7) | 39 (18.7) |
| Moderate | 38 (18.5) | 26 (12.5) |
| Severe | 94 (45.8) | 96 (46.2) |
| **Physical health self-rating, n (%)** | | |
| Some symptoms but no disease | 87 (42.4) | 85 (40.9) |
| Minor illness | 24 (11.7) | 38 (18.3) |
| Moderate to severe | 24 (11.7) | 24 (11.5) |
| **Mental health self-rating, n (%)** | | |
| Some symptoms but no disease | 68 (33.2) | 60 (28.9) |
| Minor illness | 34 (16.6) | 48 (23.1) |
| Moderate to severe | 76 (37.1) | 72 (34.6) |
| **Online mental health services search in the last 2 weeks, n (%)[d]** | | |
| Yes | 79 (38.5) | 54 (26.0) |
| Age (years), mean (SD) | 20.89 (2.32) | 21.30 (2.38) |
| Utility score, mean (SD) | 0.56 (0.26) | 0.56 (0.26) |

[a]Other includes transgender and agender participants.

[b]Employment includes paid and unpaid (volunteer) workers.

[c]K10: Kessler-10.

[d]*P*=.01.

## Service Utilization

Use of health services is reported for baseline and 1- and 3-month follow-up periods in Table 2. GP services were the most commonly utilized services for both the groups at each time point. However, the only statistically significant service between intervention and control groups was online services at baseline. A subgroup analysis indicated that the *Link* intervention was associated with a lower number of lengthy health professional consultations; however, this difference did not reach statistical significance. For example, there were less people (2 vs 11) attending extensive GP consultations (duration over 40 min) in the intervention group at the 1-month follow-up compared with those who used usual search strategies.

## Outcomes

The estimated mean AQoL-4D utility values and QALYs for the intervention and control groups over the 3-month follow-up are presented in Table 3. The utility values increased over time for the intervention group but not for the control group. At the 3-month follow-up, the estimated mean utility value for the intervention group was significantly greater than for the control group (0.63 vs 0.56, *P*<.001). Similarly, there was a statistically significant difference in QALYs at the 3-month follow-up between the groups, which favored the intervention group (0.103 vs 0.093, *P*=.01).

**Table 2.** Health service uses at baseline and 1-month and 3-month follow-ups.

| Service type[a] | Baseline, n (%) | | 1 month, n (%) | | 3 months, n (%) | |
|---|---|---|---|---|---|---|
| | Intervention | Control | Intervention | Control | Intervention | Control |
| General practitioner | 135 (65.9) | 128 (61.5) | 51 (24.9) | 53 (25.5) | 58 (28.3) | 47 (22.6) |
| Psychologist | 47 (22.9) | 56 (26.9) | 21 (10.2) | 22 (10.6) | 24 (11.7) | 22 (10.6) |
| Psychiatrist | 18 (8.8) | 27 (13.0) | 6 (2.9) | 9 (4.3) | 5 (2.4) | 7 (3.4) |
| Headspace | 23 (11.2) | 22 (10.6) | 14 (6.8) | 15 (7.2) | 21 (10.2) | 11 (5.3) |
| Other service | 16 (7.8) | 12 (5.8) | 14 (6.8) | 9 (4.3) | 13 (6.3) | 7 (3.4) |
| Online services | 79 (38.5)[b] | 54 (26.0)[b] | 52 (25.3) | 50 (24.0) | 38 (18.5) | 36 (17.3) |
| Medication | 44 (21.5) | 56 (26.9) | 19 (9.3) | 24 (11.5) | 20 (9.8) | 22 (10.6) |
| Hospital | 26 (12.7) | 24 (11.5) | 2 (1.0) | 6 (2.9) | 3 (1.5) | 4 (1.9) |
| No services used | 55 (26.8) | 57 (27.4) | 52 (25.3) | 46 (22.1) | 49 (23.9) | 57 (27.4) |

[a]Subcategories are not mutually exclusive.

[b]*P*=.01.

**Table 3.** Mean costs per participant (in Aus $) by condition cumulative over the 1- or 3-month follow-up period (based on intention-to-treat sample, N=403).

| Costs | 1-month follow-up | | | 3-month follow-up | | |
|---|---|---|---|---|---|---|
| | Intervention, mean (95% CI), Aus $ | Control, mean (95% CI), Aus $ | *P* value | Intervention, mean (95% CI), Aus $ | Control, mean (95% CI), Aus $ | *P* value |
| Consultation costs | 98 (73-123) | 161 (103-220) | .01 | 214 (148-281) | 206 (139-272) | .12 |
| Hospital costs[a] | 35 (0-94) | 10 (0-19) | —[b] | 46 (0-131) | 107 (0-305) | — |
| Medication costs | 7 (3-12) | 7 (4-10) | .29 | 16 (6-25) | 11 (6-16) | .05 |
| Total costs (health care perspective) | 145 (75-214) | 178 (119-237) | .13 | 280 (168-392) | 323 (106-540) | .64 |
| Utility | 0.60 (0.56-0.64) | 0.55 (0.51-0.59) | .17 | 0.64 (0.60-0.68) | 0.56 (0.52-0.60) | .003 |
| Quality-adjusted life years | 0.049 (0.046-0.051) | 0.047 (0.044-0.049) | .37 | 0.103 (0.097-0.109) | 0.093 (0.087-0.099) | .01 |

[a]Including inpatient and outpatient hospital costs.

[b]Insufficient observations for the 2-part model.

## Cost

As shown in Table 3, the average consultation costs at the 1-month follow-up and medication costs at the 3-month follow-up in the *Link* group were statistically significantly higher than those in the control group. No statistically significant differences for other cost categories at any other time points were found. The average total health sector costs for the intervention group were lower than the control group at 1-month and 3-month follow-ups. However, these differences were not statistically significantly different at both follow-up time points. The details of 2-part model results for medication and consultation cost are presented in Multimedia Appendix 3.

## Cost-Effectiveness

The results of the incremental analysis suggest that the *Link* intervention was associated with significantly higher utility-based quality of life than the control condition (mean difference 0.01, 95% CI 0.00-0.02). Furthermore, the *Link* intervention was also associated with lower costs (mean difference −81, 95% CI −348 to 186) compared with the control group; however, this difference did not reach statistical significance (Table 4). An intention-to-treat analysis indicated that the *Link* intervention was dominant (ie, more effective and less costly) compared with usual search strategies (95% CI dominant to Aus $11,867 per QALY).

A probabilistic analysis showed that 100% of uncertainty iterations of the ICER fell below the threshold of Aus $28,303 per QALY gained, and 73% of iterations fell in the dominant quadrant of the cost-effectiveness plane (ie, more effective and less costly, Figure 2). The cost-effectiveness acceptability curve indicated that the *Link* intervention had a 95% probability of being cost-effective as long as the threshold of willingness-to-pay is over Aus $10,000 per QALY gains (Figure 3).

**Table 4.** Results of primary and sensitivity analyses (based on 3000 bootstrap simulations).

| Analysis | Incremental costs, Aus $ (95% CI) | Incremental effects, quality-adjusted life year (95% CI) | ICER[a], mean (95% CI) | Distribution over the ICER plane (%) | | | |
|---|---|---|---|---|---|---|---|
| | | | | NE[b] | NW[b] | SE[b] | SW[b] |
| **Primary analysis** | | | | | | | |
| Intention-to-treat analysis | −79 (−342 to 134) | 0.01 (0.01 to 0.02) | Dominant (dominant to Aus $11,928) | 27 | __[c] | 73 | — |
| Complete case analysis | −130 (−590 to 226) | 0.01 (0.00 to 0.02) | Dominant (dominant to Aus $24,529) | 29 | — | 71 | — |
| **Sensitivity analysis** | | | | | | | |
| Dropout rate 10% (cover 17% population); cost development per case: Aus $3.82 | −85 (−363 to 134) | 0.01 (0.00 to 0.02) | Dominant (dominant to Aus $13,035) | 25 | — | 75 | — |
| Dropout rate 90% (cover 2% population); cost development per case: Aus $34.40 | −50 (−319 to 159) | 0.01 (0.00 to 0.02) | Dominant (dominant to Aus $14,564) | 37 | — | 63 | — |

[a]ICER: incremental cost-effectiveness ratio, based on 3000 bootstrap simulation.

[b]In the northeast (NE) quadrant, the intervention is cost-effective if the ICER falls under the specified value-for-money criterion because the intervention is more effective and costlier than the comparator. In the southeast (SE) quadrant, the intervention is less costly and more effective than the comparator (ie, dominant); therefore, the intervention is likely to be excellent for value-for-money. In the southwest (SW) quadrant, the intervention is less costly and less effective; therefore, the decision to adopt the intervention may be based on decision-makers willingness to accept some health loss relative to cost-saving. Finally, in the northwest (NW) quadrant, the results show that the intervention is associated with greater costs but less health gain, therefore, not a good option to adopt.

[c]Not applicable.

**Figure 2.** Cost-effectiveness plane of 3000 replicates of the incremental cost-effectiveness ratio—intent-to-treat analysis. QALY: quality-adjusted life year.

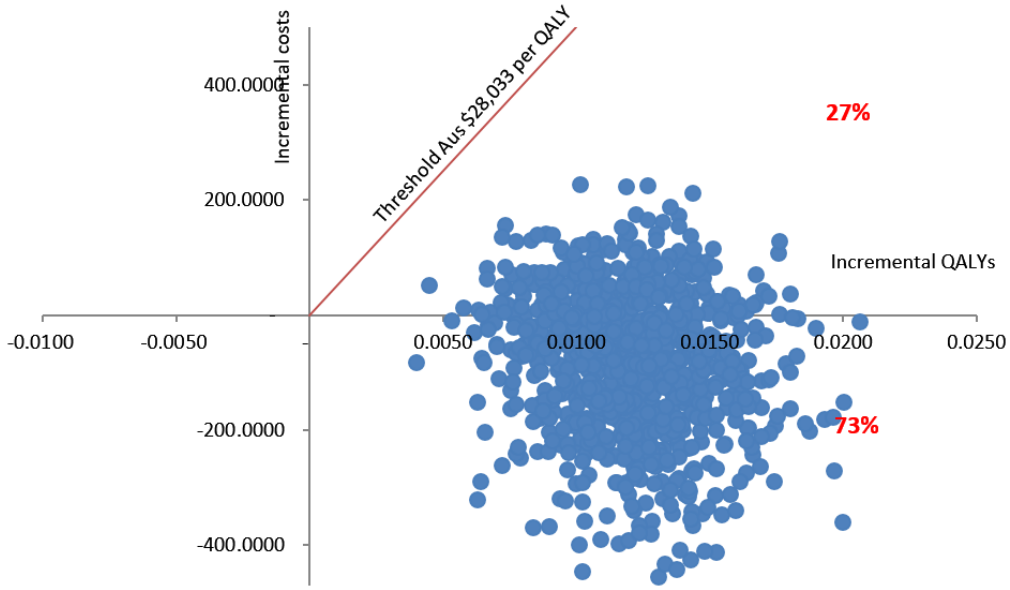
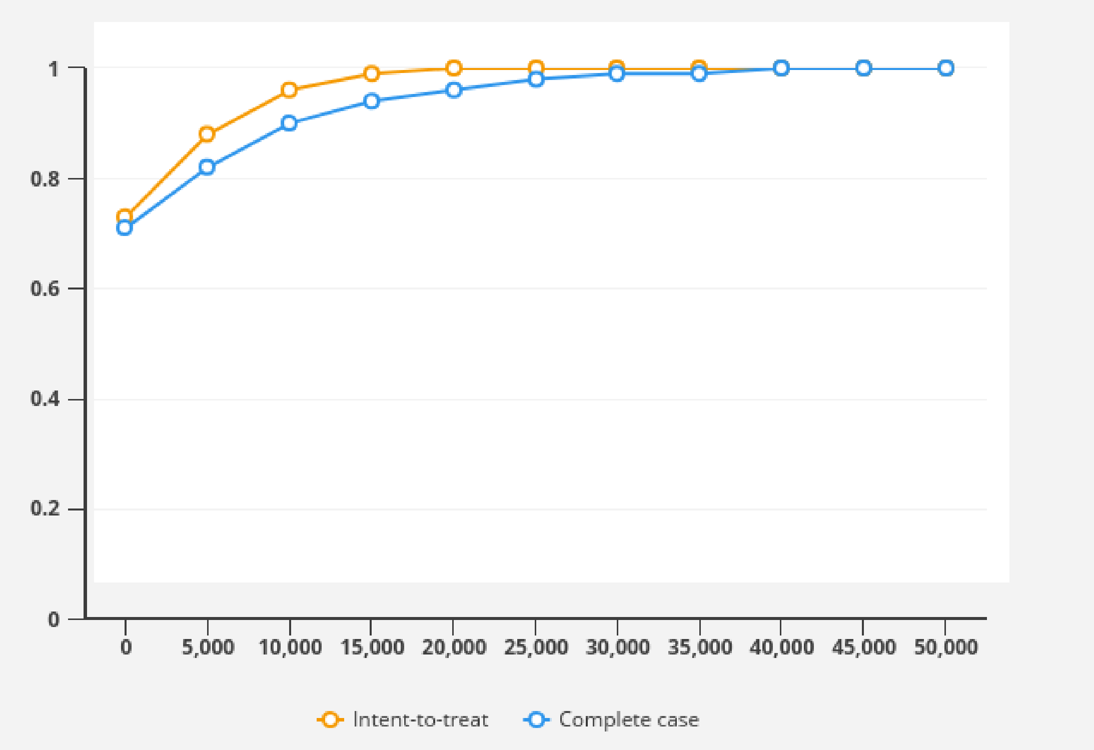


**Figure 3.** Cost-effectiveness acceptability curves for intent-to-treat and complete case analysis.




### Sensitivity Analyses

The results for the intention-to-treat (using multiple imputation) concur with those for the complete case dataset, which show a similar pattern of greater effectiveness and less cost associated with the intervention group compared with the control group (Table 4 and Multimedia Appendix 4). The sensitivity analyses, which varied the proportion of people likely to receive the intervention from 2% to 17% of young adults aged 18 to 25 years, showed that results were very robust (Table 4).

## Discussion

### Principal Findings

This study was the first cost-utility analysis of an online intervention to increase mental health help-seeking for young adults (*Link*) compared with usual search strategies. Young people randomized to the Link intervention had significantly higher utility values and QALYs gained at 3 months compared with young people using their usual online search strategies. The online help-seeking intervention was also associated with lower average total costs from a health sector perspective although this did not reach statistical significance. The online help-seeking intervention was found to be a cost-effective treatment option compared with young adults' current search strategies with a 73% probability that *Link* would be cost-saving and a 100% probability that it would be cost-effective using a willingness-to-pay threshold of Aus $28,033 per QALY gained. In fact, results suggest that even at a more modest Aus $10,000 per QALY value-for-money threshold, the intervention is still likely to be very cost-effective. The results were robust in the sensitivity analysis when complete case analysis was conducted, or the intervention costs were varied.

Interestingly, QALYs were improved in the *Link* group; however, the intervention did not appear to change resources used because quantities and costs of services were largely similar across the 2 groups. A possible explanation for this finding is that Link connected young people with higher quality, evidence-based targeted services compared with the treatment they might otherwise access.

### Comparison With Previous Work

The findings from this study are difficult to compare with other economic evaluations of internet-based interventions; as to our knowledge, this is the first study to evaluate the cost-effectiveness of an online resource to facilitate help-seeking behavior. Economic evaluations of internet interventions for mental health have been mostly focused on the treatment or prevention of mental disorders [39]. It is noteworthy that our study results are similar to economic evaluations that support the cost-effectiveness of guided internet educational and psychological interventions for the prevention and treatment of mental disorders [39,40]. More encouraging, our study indicated that the help-seeking intervention may be a very cost-effective, if not a cost-saving, option. Further research is required to confirm this result.

### Implications

Findings from our study showed that although there were no significant differences in terms of health care service use between the 2 groups, the *Link* intervention was significantly associated with lower health professional consultation costs at short-term follow-up (1 month). The reason might be that the intervention was associated with a reduction in the quantity of longer health professional (eg, GP or psychologist) consultations (duration over 40 min) than in the control group. Another important point is that the *Link* intervention was associated with lower hospitalization costs than the control at the 3-month follow-up, although the number of people who were hospitalized was not different. This might suggest a positive benefit of the *Link* intervention in reducing severity of mental health problems that require intensive treatments. However, these results did not reach significance, given that the sample sizes of these subgroups were relatively small. As noted above, these results may be explained by the quality of services being accessed via the *Link* platform. Further research with larger sample sizes and perhaps more evaluation of the type of care being accessed (in terms of quality) is needed.

This study, for the first time, raises the possibility that improving help-seeking not only assists young adults in accessing mental health care services but is also associated with quality of life improvements. More importantly, a Web-based mental health service navigation website (ie, *Link* platform) demonstrated a high probability of being cost-effective. The initial results from this study are certainly very promising and suggest that if access to the intervention was increased, this could result in significant health impacts and likely cost savings.

### Strengths and Limitations

This study has several strengths. First, this study used a cost-utility framework whereby outcomes are expressed as QALYs [37], thereby allowing results to be comparable with other economic evaluations and commonly used value-for-money thresholds. Second, this study adheres to the CHEERS checklist, which are quality reporting guidelines for economic evaluation [13]. Finally, a sensitivity analysis has been conducted to assess the robustness of the findings from the primary analysis.

In terms of limitations, these results do not include any costs beyond the health sector, which may underestimate the cost-effectiveness of the *Link* intervention. For example, the inclusion of productivity costs (absenteeism and presentism) may be associated with even more cost savings. The study was also limited by the relatively short time horizon (ie, 3 months) and the use of self-reported retrospective utilization of health care services and medication, potentially leading to recall bias. It is not clear whether this may have led to an over- or underestimation of resource use reporting, although any biases are likely to be the same in both groups. Further research using a broader societal perspective and longer follow-up is needed.

### Conclusions

In conclusion, the online help-seeking navigation website, *Link*, appears to provide a cost-effective and, possibly, cost-saving tool for young adults compared with the usual methods for seeking care. The intervention demonstrated a reduction in health care professional consultation costs at the 1-month follow-up and hospital costs at the 3-month follow-up. These results were robust in the sensitivity analysis. Further research to confirm these results could have important implications for increasing the accessibility of mental health care services for young adults.


## Acknowledgments

This study was funded by the Young and Well Cooperative Research Centre, an Australian-based international research center that unites young people with researchers, practitioners, innovators, and policymakers from over 70 partner organizations. SC is supported by a National Health and Medical Research Council (NHMRC) Career Development Fellowship. During the conduct of this work, CM was funded by a NHMRC Early Career Fellowship Grant (APP1035887).


## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Consolidated Health Economic Evaluation Reporting Standards Statement checklist.

[DOCX File, 18KB - jmir_v21i7e13065_app1.docx ]

## Multimedia Appendix 2

Unit costs for health care consultation sourced from the 2014 Medicare Benefit Schedule Book.

[DOCX File, 13KB - jmir_v21i7e13065_app2.docx ]

## Multimedia Appendix 3

Two-part models for consultation and medication costs (intent-to-treat analysis).

[DOCX File, 15KB - jmir_v21i7e13065_app3.docx ]

### Multimedia Appendix 4

Mean costs per participant (in Aus $) by condition cumulative over the 1- or 3-month follow-up period (based on completer analysis).

[DOCX File, 15KB - jmir_v21i7e13065_app4.docx ]

### Multimedia Appendix 5

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2MB - jmir_v21i7e13065_app5.pdf ]

## References


1. Erskine HE, Moffitt TE, Copeland WE, Costello EJ, Ferrari AJ, Patton G, et al. A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. Psychol Med 2015 May;45(7):1551-1563 [FREE Full text] [doi: 10.1017/S0033291714002888] [Medline: 25534496]
2. Lee YC, Chatterton ML, Magnus A, Mohebbi M, Le LK, Mihalopoulos C. Cost of high prevalence mental disorders:findings from the 2007 Australian national survey of mental health and wellbeing. Aust N Z J Psychiatry 2017 Dec;51(12):1198-1211. [doi: 10.1177/0004867417710730] [Medline: 28565923]
3. Burgess PM, Pirkis JE, Slade TN, Johnston AK, Meadows GN, Gunn JM. Service use for mental health problems: findings from the 2007 national survey of mental health and wellbeing. Aust N Z J Psychiatry 2009 Jul;43(7):615-623. [doi: 10.1080/00048670902970858] [Medline: 19530018]
4. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
5. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113. [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
6. Ali K, Farrer L, Fassnacht DB, Gulliver A, Bauer S, Griffiths KM. Perceived barriers and facilitators towards help-seeking for eating disorders: a systematic review. Int J Eat Disord 2017 Jan;50(1):9-21. [doi: 10.1002/eat.22598] [Medline: 27526643]
7. Anderson KK, Fuhrer R, Malla AK. 'There are too many steps before you get to where you need to be': help-seeking by patients with first-episode psychosis. J Ment Health 2013 Aug;22(4):384-395. [doi: 10.3109/09638237.2012.705922] [Medline: 22958140]
8. Tylee A, Haller DM, Graham T, Churchill R, Sanci LA. Youth-friendly primary-care services: how are we doing and what more needs to be done? Lancet 2007 May 5;369(9572):1565-1573. [doi: 10.1016/S0140-6736(07)60371-7] [Medline: 17482988]
9. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
10. Lingley-Pottie P, McGrath PJ, Andreou P. Barriers to mental health care: perceived delivery system differences. ANS Adv Nurs Sci 2013;36(1):51-61. [doi: 10.1097/ANS.0b013e31828077eb] [Medline: 23370500]
11. Christensen H, Hickie IB. Using e-health applications to deliver new mental health services. Med J Aust 2010 Jun 7;192(11 Suppl):S53-S56. [Medline: 20528711]
12. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 4;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
13. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. Consolidated health economic evaluation reporting standards (CHEERS)--explanation and elaboration: a report of the ISPOR health economic evaluation publication guidelines good reporting practices task force. Value Health 2013;16(2):231-250 [FREE Full text] [doi: 10.1016/j.jval.2013.02.002] [Medline: 23538175]
14. Paul D, Nolan K, Koller C. The University of Newcastle. 2013. QuON User Help Files URL: https://quon.newcastle.edu.au/quon_doc/
15. Sanci L, Kauer S, Thuraisingam S, Chondros P, Davidson S, Duncan AM, et al. Positive affect and help-seeking effects of a mental health service navigation website for young adults called Link: a randomised controlled trial. JMIR Ment Health 2019:- (forthcoming). [doi: 10.2196/preprints.13189]
16. 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]
17. Rickwood D, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. Aust E J Adv Mental Health 2005;4(3):218-251. [doi: 10.5172/jamh.4.3.218]
18. Kauer S, Buhagiar K, Sanci L. Facilitating mental health help seeking in young adults: the underlying theory and development of an online navigation tool. Adv Ment Health 2016 Oct 31;15(1):71-87 [FREE Full text] [doi: 10.1080/18387357.2016.1237856]

19. Kauer SD, Buhagiar K, Blake V, Cotton S, Sanci L. Facilitating mental health help-seeking by young adults with a dedicated online program: a feasibility study of Link. BMJ Open 2017 Jul 9;7(7):e015303 [FREE Full text] [doi: 10.1136/bmjopen-2016-015303] [Medline: 28694345]
20. Hawthorne G, Richardson J, Osborne RH. The assessment of quality of life (AQoL) instrument: a psychometric measure of health-related quality of life. Qual Life Res 1999;8(3):209-224. [doi: 10.1023/A:1008815005736]
21. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials. Oxford: OUP Oxford; 2014.
22. The People of Australia: Statistics From the 2011 Census. Canberra: Australian Bureau of Statistics; 2015.
23. Australian Bureau of Statistics. 2015. The 2014-15 National Health Survey: Psychological Distress URL: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4364.0.55.001~2014-15~Main%20Features~Psychological%20distress~16
24. Cave L, Fildes J, Luckett G, Wearring A. Mission Australia. 2015. Youth Survey Report 2015 URL: https://www.missionaustralia.com.au/publications/youth-survey/413-mission-australia-youth-survey-2015/file
25. Chatterton ML, Mihalopoulos C, O'Neil A, Itsiopoulos C, Opie R, Castle D, et al. Economic evaluation of a dietary intervention for adults with major depression (the 'SMILES' trial). BMC Public Health 2018 Dec 22;18(1):599 [FREE Full text] [doi: 10.1186/s12889-018-5504-8] [Medline: 29783983]
26. Department Human Services, Australian Government. 2015. Pharmaceutical Benefits Schedule Item Reports URL: http://medicarestatistics.humanservices.gov.au/statistics/pbs_item.jsp
27. Independent Hospital Pricing Authority. 2014. National Efficient Price Determination 2014-15 URL: https://www.ihpa.gov.au/sites/g/files/net636/f/publications/national_efficient_price_determination_2014-15.pdf
28. Australian Psychological Society. 2015. APS National Schedule of Recommended Fees (not including GST) and Item Numbers* for Psychological Services.
29. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci 2007 Sep;8(3):206-213. [doi: 10.1007/s11121-007-0070-9] [Medline: 17549635]
30. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
31. Belotti F, Deb P, Manning WG, Norton EC. Twopm: two-part models. Stata J 2015;15(1):3-20. [doi: 10.1177/1536867X1501500102]
32. Deb P, Norton EC. Modeling health care expenditures and use. Annu Rev Public Health 2018 Dec 1;39:489-505. [doi: 10.1146/annurev-publhealth-040617-013517] [Medline: 29328879]
33. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-an ISPOR good research practices task force report. Value Health 2015 Mar;18(2):161-172 [FREE Full text] [doi: 10.1016/j.jval.2015.02.001] [Medline: 25773551]
34. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics 2014 Dec;32(12):1157-1170 [FREE Full text] [doi: 10.1007/s40273-014-0193-3] [Medline: 25069632]
35. Briggs AH, Gray AM. Handling uncertainty in economic evaluations of healthcare interventions. Br Med J 1999 Sep 4;319(7210):635-638 [FREE Full text] [doi: 10.1136/bmj.319.7210.635] [Medline: 10473486]
36. Edney LC, Haji AA, Cheng TC, Karnon J. Estimating the reference incremental cost-effectiveness ratio for the Australian health system. Pharmacoeconomics 2018 Dec;36(2):239-252. [doi: 10.1007/s40273-017-0585-2] [Medline: 29273843]
37. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. Fourth Edition. Oxford: Oxford University Press; 2015.
38. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annu Rev Public Health 2002;23:151-169. [doi: 10.1146/annurev.publhealth.23.100901.140546] [Medline: 11910059]
39. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
40. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]


## Abbreviations

**AQoL:** Assessment of Quality of Life
**AR-DRG:** Australian Refined Diagnostic Related Group
**CHEERS:** Consolidated Health Economic Evaluation Reporting Standards Statement
**GLM:** general linear model
**GP:** general practitioner
**ICER:** incremental cost-effectiveness ratio
**K10:** Kessler-10

**NHMRC:** National Health and Medical Research Council
**QALY:** quality-adjusted life year
**RCT:** randomized controlled trial
**RUQ:** resource use questionnaire

*Edited by G Eysenbach; submitted 19.12.18; peer-reviewed by C Buntrock, F Kählke; comments to author 14.02.19; revised version received 10.04.19; accepted 07.05.19; published 22.07.19.*

*Please cite as:*
*Le LKD, Sanci L, Chatterton ML, Kauer S, Buhagiar K, Mihalopoulos C*
*The Cost-Effectiveness of an Internet Intervention to Facilitate Mental Health Help-Seeking by Young Adults: Randomized Controlled Trial*
*J Med Internet Res 2019;21(7):e13065*
*URL: http://www.jmir.org/2019/7/e13065/*
*doi:10.2196/13065*
*PMID:31333199*

Review

# An Evaluation of the Effectiveness of the Modalities Used to Deliver Electronic Health Interventions for Chronic Pain: Systematic Review With Network Meta-Analysis

Brian W Slattery[1,2], PhD; Stephanie Haugh[2], MSc; Laura O'Connor[2], MIT; Kady Francis[2], MSc; Christopher P Dwyer[2], PhD; Siobhán O'Higgins[2], PhD; Jonathan Egan[2], PsyD; Brian E McGuire[2], PhD

[1]School of Nursing and Human Sciences, Dublin City University, Dublin, Ireland

[2]Centre for Pain Research, School of Psychology, National University of Ireland, Galway, Ireland

**Corresponding Author:**
Brian W Slattery, PhD
School of Nursing and Human Sciences
Dublin City University
Glasnevin
Dublin, 9
Ireland
Phone: 353 017008034
Email: brian.slattery@dcu.ie

## Abstract

**Background:** Electronic health (eHealth) is the use of information and communication technology in the context of health care and health research. Recently, there has been a rise in the number of eHealth modalities and the frequency with which they are used to deliver technology-assisted self-management interventions for people living with chronic pain. However, there has been little or no research directly comparing these eHealth modalities.

**Objective:** The aim of this systematic review with a network meta-analysis (NMA) is to compare the effectiveness of eHealth modalities in the context of chronic pain.

**Methods:** Randomized controlled trials (N>20 per arm) that investigated interventions for adults with chronic pain, delivered via an eHealth modality, were included. Included studies were categorized into their primary node of delivery. Data were extracted on the primary outcome, pain interference, and secondary outcomes, pain severity, psychological distress, and health-related quality of life. Pairwise meta-analyses were undertaken where possible, and an NMA was conducted to generate indirect comparisons and rankings of modalities for reducing pain interference.

**Results:** The search returned 18,470 studies with 18,349 being excluded (duplicates=2310; title and abstract=16,039). Of the remaining papers, 30 studies with 5394 randomized participants were included in the review. Rankings tentatively indicated that modern eHealth modalities are the most effective, with a 43% chance that mobile apps delivered the most effective interventions, followed by a 34% chance that interventions delivered via virtual reality were the most effective.

**Conclusions:** This systematic review with an NMA generated comparisons between eHealth modalities previously not compared to determine which delivered the most effective interventions for the reduction of pain interference in chronic pain patients. There are limitations with this review, in particular, the underrepresented nature of some eHealth modalities included in the analysis. However, in the event that the review is regularly updated, a clear ranking of eHealth modalities for the reduction of pain interference will emerge.

***(J Med Internet Res 2019;21(7):e11086)*** doi:10.2196/11086

**KEYWORDS**

eHealth; mHealth; digital health; Virtual Reality; chronic pain; systematic review; network meta-analysis

## Introduction

### Electronic Health

As technological advances pervade every aspect of daily life, there has been a corresponding proliferation in the development and implementation of technological interventions for health-related purposes. Electronic health (eHealth), the broad term for information and communication technologies deployed in health settings, is a growing area of interest as the international research community attempts to address issues facing modern health care [1]. Typically, an eHealth *modality* is considered to be some specific form of technology that is applied in the context of health care [2-4]. Examples of eHealth modalities include internet-based (Web-based health interventions [5-9], telephone-supported (interventions with telephone support from health practitioners) [10], interactive voice response (the use of a phone's touch-tone keypad to provide responses to automated scripts) [11,12], virtual reality (a 3-dimensional computer-generated environment that the individual can explore, interact with, and manipulate) [13,14], videoconferencing (the use of high-quality real-time video and audio connection via online internet networks) [15], and mobile phone apps (mobile-based or mobile-enhanced programs) that deliver health-related services [16,17]. More detailed definitions of the various types of eHealth interventions are available in Multimedia Appendix 1.

The core value proposition for delivering health care via an eHealth modality is that the barriers experienced by traditional in-person treatment methods are reduced or potentially removed [4,18-21]. For instance, a Web-based eHealth intervention may improve accessibility to treatment, reduce the waiting list duration, and can be delivered more cost-effectively than in-person services [22]. For these reasons, eHealth has gained considerable traction for conditions that are long-term and where there is a shift toward self-management [23-26]. In this context, where ongoing disease management is required, eHealth interventions offer a viable and important support option. Many eHealth solutions have been developed for a variety of chronic illnesses, including diabetes [27,28], breast cancer [29], hypertension [30], cardiovascular disease [16], multiple sclerosis [31], headache [8], and chronic pain [12,14,32-38].

### Electronic Health and Chronic Pain

Chronic pain refers to pain that lasts for more than 3 months [39]. Chronic pain encompasses many diverse conditions, is highly prevalent, and is a leading cause of long-term disability [39]. Much eHealth research has been conducted in the area of chronic (noncancer) pain, and eHealth interventions have shown to be efficacious in reducing pain interference [40]. However, despite the increasing variety of eHealth modalities used for chronic pain, studies typically focus on 1 modality, and as a result, direct comparisons of modalities are rare [22]. Identifying the need to investigate the relative strengths and weaknesses of modality types, Heapy et al conducted a systematic review of eHealth self-management interventions for chronic pain, in which three modality types were evaluated, namely, telephone, interactive voice response, and internet. They concluded that each modality was effective in the context of chronic pain, but no conclusive evidence points to one being more superior than the others.

Notably, Heapy et al began the necessary steps toward ascertaining the varying efficacies of each modality as the *contributing factor* to intervention success. However, the authors recognized certain limitations with their review, such as the breadth of their search strategy (ie, limited to three databases) and the low number (ie, 3) of included eHealth modalities. Moreover, the review included a variety of study designs, and although they reported on the between-condition effect sizes when possible, a quantitative comparison (ie, a meta-analysis) was not conducted. Therefore, one of Heapy et al's indications for future research was to identify the relative efficacy of modality types through direct comparison.

### Why Is It Important to Do This Review?

Although there are a growing number of eHealth interventions for chronic pain, there is a stark lack of research comparing eHealth modalities in this context. Directly comparing eHealth modalities deployed in chronic pain research could potentially yield important insights into which modalities are more efficacious in what context and for what reasons (eg, treatment fidelity, resource availability, issues with target population, typical engagement levels, and cost efficiency). Thus, from the perspectives of patient well-being, health care provision, and optimizing research interventions, there is an impetus to first identify the most effective modalities for chronic pain and to then investigate why they are the most effective.

The aim of this study was to add to the literature that concerns itself with evaluating eHealth modalities in the context of reducing pain interference for chronic pain patients by directly comparing treatment outcomes across studies that have deployed an eHealth modality. Critically, this review conducted a network meta-analysis (NMA) and quantitatively compared and ranked the eHealth modalities used for interventions in chronic pain, which has not been done before. An NMA is an extension of a meta-analysis and enables multiple treatments to be compared using direct and indirect comparisons across trials using a common comparator [41-43].

### Objective

The objective of this study was to conduct a systematic review and an NMA to evaluate and compare the effectiveness of the eHealth modalities used to deliver interventions (other than drugs) for adults living with chronic noncancer pain.

## Methods

### Protocol and Registration

The systematic review and NMA were conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and the PRISMA Network Meta-Analysis extension statement (see Multimedia Appendix 2) [44]. The protocol for this study is registered with the International Prospective Register of Systematic Reviews database (registration number: CRD42016035595) [45].

## Outcomes

### *Primary Outcome*

Similar to previous research [33,34,46,47], and in accordance with outcome measures outlined by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials [48], pain interference was the primary outcome variable. Where pain interference was not reported, pain-related disability or a reverse-scored measure of physical functioning was extracted.

### *Secondary Outcomes*

Secondary outcomes were measures of pain severity, psychological distress (measures of depression were extracted where available; measures of anxiety and reverse-scored measures of mental health were also acceptable), and health-related quality of life (HRQoL).

## Eligibility Criteria

The eligibility requirements for included studies in this review are outlined in Table 1. All studies included in this review were required to be published in peer-reviewed journals and available in English. The criteria were influenced by a Cochrane review of internet-delivered psychological therapies for chronic pain by Eccleston et al [49].

## Classification

Studies were merged to create nodes representing the primary delivery method (eg, internet). A study was not included in the network if both arms were classified as the same modality without an additional comparator.

## Information Sources

A total of 4 databases, Cochrane Central Register of Controlled Trials (CENTRAL; Cochrane Library), Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica dataBASE (EMBASE), and PsycINFO, were searched from inception until November 22, 2017. Necessary changes were made to adapt the search terms for different interfaces. The search strategy is detailed in Textbox 1.

The reference lists of relevant systematic reviews and of included studies were screened to identify any relevant studies. The metaRegister of Controlled Trials [50], Clinicaltrials.gov [51], and the World Health Organization's International Clinical Trials Registry Platform [52] were also searched.

## Study Selection

Members of the research team screened titles and abstracts to search for duplicate and nonrelevant studies; 10% of the papers were assessed in duplicate. In total, 2 review authors (SH and KF) independently screened full-text papers for inclusion. Studies were included if they (1) were randomized controlled trials (RCTs); (2) had N>20 per arm at each time point; (3) had participants with noncancer-related chronic pain; (4) were delivered via eHealth modality; and (5) measured a suitable pain outcome.

**Table 1.** Eligibility criteria (Population, Intervention, Comparison, Outcome, Study Design) included in this review.

| Category | Eligibility criteria |
|---|---|
| Population | Adults with noncancer-related chronic pain |
| Intervention | Interventions for managing chronic pain delivered via an electronic health (eHealth) modality |
| Comparison intervention | At least one of the following: an active eHealth intervention; enhanced control; treatment-as-usual; waiting-list control |
| Outcome measures | Pain interference; pain severity; psychological distress; health-related quality of life |
| Study design | Randomized controlled trials |

**Textbox 1.** Search terms.

1. (Telecommunications)/ OR (telemedicine OR tele-medicine).mp OR (telehealth OR tele-health).mp OR (ehealth OR e-health).mp OR (mobile health OR mhealth OR m-health).mp OR (ICT).mp OR ((inform* OR communicat* OR interact*) adj6 (computer* OR technolog* OR software)).mp OR ((health* OR treat* OR therap* or intervention* OR assist* OR selfmanag* OR self-manag*) adj6 (computer* OR technolog* OR software)).mp OR (internet)/ OR (internet* OR world wide web OR www OR web-based OR email OR e-mail OR online).mp OR (telephone* OR phone* OR mobile* OR cellphone* OR cellular telephone* OR application* OR app* OR text* OR SMS OR smartphone* OR mobile operating system technolog* OR microcomputer* ).mp OR (virtual reality OR augmented reality OR VR OR AR).mp OR (IVR OR interactive voice response OR voice response unit OR VRU OR speech recognition OR voice recognition).mp

   AND

2. (Pain)/ OR (Pain Measurement)/ OR (Headache disorders)/ OR (Fibromyalgia)/ OR (pain* OR headache* OR migraine* OR fibromyalgia* OR neuralgia*).mp OR (pain intensity OR pain severity OR pain outcome*) OR (self-reported pain)

   AND

3. "Chronic pain" OR headache*

   AND

4. (randomized controlled trial OR randomised controlled trial.pt) OR (controlled clinical trial.pt) OR (randomized.ab OR randomised.ab) OR (placebo.ab) OR (clinical trials as topic.sh) OR (randomly.ab) OR (trial.ti) OR (groups.ti)

### Data Collection Process and Data Items

Data were independently extracted by 2 authors (BS and SH) into a preprepared excel sheet. The following items were extracted: means and SDs at postintervention for pain interference, psychological distress and HRQoL, sample size, measures, mean age, percentage of females, diagnosis, mean years of pain, method of recruitment, and presence of contact with researchers or therapists. If no SDs were reported, they were calculated from the available SEs or CIs.

### Risk of Bias

In line with previous research, risk of bias within individual studies was assessed using the Cochrane Risk of Bias tool. Please see the published protocol for additional details [1]. Funnel plots and Egger tests were conducted to investigate publication bias across studies.

### Geometry of the Network

The network includes a node for each eHealth modality. In addition, the network contains both a control node (comprised wait list control and treatment-as-usual control groups) and an enhanced control node (eg, educational booklet).

### Summary Measures

Standardized mean differences (SMDs) between groups at postintervention and measures of uncertainty are reported. Additional summary measures such as treatment rankings and the probability of each modality arm being the best are reported.

### Planned Methods of Analysis

Random-effects pairwise meta-analyses of each available comparison were run as an exploratory analysis using Stata 13 (StataCorp LLC). These analyses were carried out on both the primary and secondary outcomes: pain interference, pain severity, psychological distress, and HRQoL.

An NMA random-effects model of the eHealth modalities used to deliver chronic pain interventions with the purpose of reducing pain interference was developed in WinBUGS 14 (MRC and Imperial College of Science, Technology and Medicine). This model was based on a Bayesian framework but was created with vague priors. The NMA returned pairwise comparisons between all modalities, rankings of the modalities and assessed the probability that each modality is the best. Tests of design inconsistency [53] and loop inconsistency [54] were run using Stata 13. Node splitting was conducted on comparisons with both direct and indirect evidence [55]. Additional information is provided in the protocol [1].

### Additional Analyses

As outlined in the protocol [1], the purpose of adding study-level covariates was to reduce heterogeneity by allowing the NMA to take account of additional information and minimize the differences between the studies within each modality. Covariates would be added to the model based on a reduction in the deviance information criterion (DIC). In this network, the added covariates did not have a significant effect. Sensitivity analyses investigating the influence of priors, initial values, length of burn-in, and testing convergence were carried out.

## Results

### Study Selection

The search returned 18,470 studies (Figure 1): PsycINFO (n=1913), MEDLINE (n=5286), EMBASE (n=10,479), and CENTRAL (n=792). There were 2310 studies that were excluded as duplicates and 16,039 studies excluded on the basis of title and abstract. In total, 122 potentially eligible studies were identified and then assessed on the basis of full text. Of these, 92 studies were excluded: 51 studies were not an RCT; 13 studies had less than 20 participants per arm at each time point; 7 studies had patients with cancer-related chronic pain; 12 studies did not deliver the intervention via an eHealth modality; 3 studies did not measure an appropriate pain outcome; and 6 studies consisted of 2 arms within the same node without an additional comparator. There were 30 studies that were included in the analysis.

### Study Characteristics

30 studies were included in this review. Each study arm was classified by the primary delivery method (eg, internet). Although the majority of intervention arms were compared with control arms, 1 study involved the comparison of 2 active treatments [56]. The 30 studies encompassed 61 arms: 23 internet-delivered arms [5-9,15,34,36,37,46,57-69]; 2 telephone [35,70]; 1 mobile app [71]; 2 virtual reality [14,72]; 1 videoconferencing [15]; 1 interactive voice response [12]; 25 control; and 13 enhanced controls (Table 2).

**Figure 1.** Flow diagram of studies assessed for eligibility.

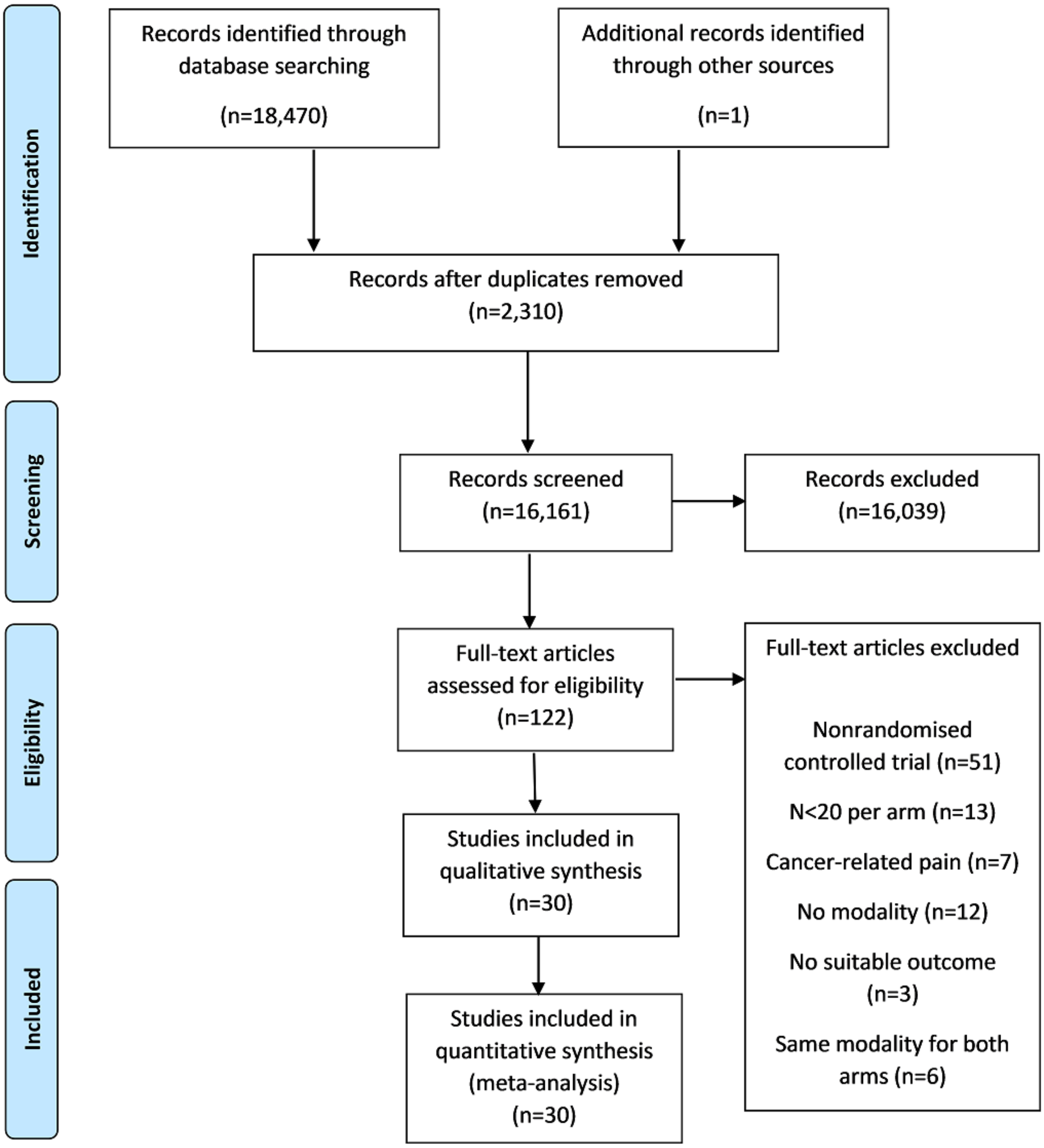


**Table 2.** Studies included in the review.

| Study | Comparison | N[a] | Pain conditions / location | Average age (years) | Gender, female, n (%) | Attrition, n (%) |
|---|---|---|---|---|---|---|
| Berman (2009) | Internet (mind-body) versus control (WLC[b]) | 89 (52, 37) | Nonspecific chronic pain | 65.8 | 68 (87.2) | 11 (12.4) |
| de Boer (2014) | Internet (CBT[c]) versus enhanced control (face-to-face CBT) | 63 (33, 30) | Nonspecific chronic pain | 52.1 | 32 (64) | 13 (20.6) |
| Bromberg (2012) | Internet versus control (TAU[d]) | 185 (92, 93) | Migraine or headache | 42.6 | 165 (89) | 19 (10.2) |
| Buhrman (2004) | Internet (CBT) versus control (WLC) | 56 (22, 29) | Chronic back pain | 44.6 | 35 (62.5) | 5 (8.9) |
| Buhrman (2011) | Internet (CBT) versus control (WLC) | 54 (26, 28) | Chronic back pain | 43.2 | 37 (68.5) | 4 (7.4%) |
| Carpenter (2012) | Internet (CBT) versus control (WLC) | 141 (70, 71) | Chronic lower back pain | 42.5 | 117 (83) | 23 (16.3) |
| Chiauzzi (2010) | Internet (self-management) versus enhanced control (text-based material) | 199 (95, 104) | Chronic back pain | 46.1 | 134 (67.7) | 15 (7.5) |
| Dear (2013) | Internet (CBT) versus control (WLC) | 62 (31, 31) | Multiple pain conditions/sites | 49 | 53 (85) | 2 (3.2) |
| Dear (2015) | Internet (CBT) versus control (WLC) | 472 (397, 75) | Multiple pain conditions/sites | 50 | 375 (80) | 50 (10.6) |
| Dear (2017) | Internet (CBT) versus enhanced control (workbook) | 164 (76, 88) | Multiple pain conditions/sites | 47.8 | 135 (82) | 14 (8.5) |
| Devineni (2005) | Internet versus control (WLC) | 86 (39, 47) | Migraine or headache | 41.3 | 111 (79.6) | 53 (38.1) |
| Garcia-Palacios (2015) | Virtual reality (activity management) versus control (TAU) | 61 (31, 30) | Fibromyalgia syndrome | 50.5 | 61 (100) | 2 (3) |
| Herbert (2017) | Videoconferencing (ACT[e]) versus enhanced control (face-to-face ACT) | 129 (65, 64) | Multiple pain conditions/sites | 52 | 23 (17.8) | 28 (21.7) |
| Kleiboer (2014) | Internet versus control (WLC) | 368 (195, 173) | Migraine or headache | 43.6 | 314 (85) | 96 (26) |
| Krein (2013) | Internet (pedometer) versus enhanced control (pedometer) | 229 (111, 118) | Chronic low back pain | 51.6 | 29 (12.7) | 22 (9.6) |
| Kristjánsdóttir (2013) | Mobile app (CBT) versus internet (CBT) | 140 (70, 70) | Chronic widespread pain | 44.2 | 140 (100) | 40 (28.6) |
| Kroenke (2014) | Telephone (care management) versus control (TAU) | 250 (124, 126) | Chronic musculoskeletal and chronic generalised pain | 55.2 | 43 (17.2) | 12 (4.8) |
| Leveille (2009) | Internet (health coaching) versus enhanced control (general health information) | 241 (121, 120) | Chronic musculoskeletal pain | 52.4 | 138 (57.3) | 99 (41.1) |
| Lin (2017) | Internet (ACT) versus control (WLC) | 302 (201, 101) | Multiple pain conditions/sites | 51.7 | 254 (84.1) | 73 (24.2) |
| Lorig (2008) | Internet (pain management) versus control (TAU) | 855 (433, 422) | Arthritis or fibromyalgia | 52.4 | 780 (91.2) | 214 (25) |
| McBeth (2012) | Telephone (CBT) versus enhanced control (exercise) versus control (TAU) | 442 (224, 109, 109) | Chronic widespread pain | 56.2 | 307 (69.5) | 81 (18.3) |
| Müller (2016) | Internet (positive psychology) versus control (text-based materials) | 96 (51, 45) | Multiple pain conditions/sites | 59.4 | 67 (69.8) | 19 (19.8) |
| Naylor (2008) | Interactive voice response (CBT) versus control (TAU) | 51 (26, 25) | Chronic musculoskeletal pain | 46 | 44 (86) | 0 (0) |
| Peters (2017) | Internet (positive psychology) versus control (WLC) | 284 (233, 51) | Multiple pain conditions/sites | 48.9 | 234 (84.7) | 70 (24.6) |

| Study | Comparison | N[a] | Pain conditions / location | Average age (years) | Gender, female, n (%) | Attrition, n (%) |
|---|---|---|---|---|---|---|
| Ruehlman (2012) | Internet (CBT) versus control (WLC) | 305 (162, 143) | Multiple pain conditions/sites | 44.9 | 195(64) | 64 (20.9) |
| Ström (2000) | Internet (applied relaxation) versus control (WLC) | 102 (20, 25) | Headache related pain | 36.7 | 69 (67.6) | 57 (56) |
| Trompetter (2015) | Internet (ACT) versus control (WLC) | 238 (161, 77) | Multiple pain conditions/sites | 52.8 | 181 (76) | 66 (27.7) |
| Williams (2010) | Internet (self-management) versus control (TAU) | 118 (59, 59) | Fibromyalgia | 50.5 | 112 (95) | 12 (10.2) |
| Wilson (2015) | Internet (pain management) versus control (WLC) | 114 (57, 57) | Chronic non-cancer pain | 49.3 | 72 (78) | 34 (29.8) |
| Yilmaz Yelvar (2017) | Virtual reality (physiotherapy) versus enhanced control (physiotherapy) | 46 (23, 23) | Non-specific low-back pain | 49.6 | 28 (63.63) | 2 (4.3) |

[a]Total N randomized (Arm 1 N, Arm 2 N, Arm 3 N [where applicable]).

[b]WLC: waitlist control

[c]CBT: cognitive behavioral therapy

[d]TAU: treatment-as-usual.

[e]ACT: acceptance and commitment therapy.

A total of 5288 participants were included in the review. There were 3005 participants randomized to interventions delivered via an eHealth modality: internet (n=2509); telephone (n=305); videoconferencing (n=65); virtual reality (n=53); mobile apps (n=47); and interactive voice response (n=26).

### Risk of Bias Within Studies

The risk of bias summary is presented in Table 3. In total, 18 studies were considered to have been effectively randomized, 11 studies did not provide adequate information, and 1 study did not describe randomization and was judged to be at a high risk of bias. Furthermore, 10 studies used appropriate methods of allocation concealment, 18 studies did not appropriately describe their allocation methods, and 2 studies were judged as high risk, given that allocation was not blinded from research assistants. A total of 23 studies were not at risk of detection bias; the majority of these administered their assessments online. Furthermore, 7 studies were considered unclear. Although 15 studies provided clear information on their levels of attrition, 14 of them were judged to be unclear, with many failing to report differences between completers and noncompleters, and 1 study was considered to be at high risk of bias because of statistical differences between the completers and noncompleters. In total, 28 studies reported all outcomes and were free from selective reporting bias. In addition, 2 studies were judged to be of high risk because data could not be extracted. No other sources of bias were found for the 30 studies.

**Table 3.** Assessment of within-study bias.

| Study | Adequate sequence generation | Allocation concealment | Blinding | Incomplete outcome data addressed | Free of selective reporting | Free of other bias |
|---|---|---|---|---|---|---|
| Berman (2009) | +[a] | –[b] | + | + | + | + |
| de Boer (2014) | + | + | + | + | + | + |
| Bromberg (2012) | + | ?[c] | + | + | + | + |
| Buhrman (2004) | + | ? | + | + | + | + |
| Buhrman (2011) | ? | + | + | + | + | + |
| Carpenter (2012) | + | ? | + | ? | – | + |
| Chiauzzi (2010) | ? | ? | ? | ? | + | + |
| Dear (2013) | – | – | + | ? | + | + |
| Dear (2015) | + | + | + | ? | + | + |
| Dear (2017) | + | + | + | ? | + | + |
| Devineni (2005) | + | ? | ? | + | + | + |
| Garcia-Palacios (2015) | + | ? | ? | ? | + | + |
| Herbert (2017) | ? | + | + | + | + | + |
| Kleiboer (2014) | + | + | + | + | + | + |
| Krein (2013) | + | ? | ? | ? | + | + |
| Kristjánsdóttir (2013) | + | ? | ? | + | + | + |
| Kroenke (2014) | + | + | + | ? | + | + |
| Leveille (2009) | ? | ? | + | + | + | + |
| Lin (2017) | + | ? | + | ? | + | + |
| Lorig (2008) | ? | ? | + | – | + | + |
| McBeth (2012) | + | + | + | + | + | + |
| Muller (2016) | + | ? | + | ? | + | + |
| Naylor (2008) | ? | ? | + | ? | + | + |
| Peters (2017) | ? | ? | + | + | + | + |
| Ruehlman (2012) | ? | ? | + | + | + | + |
| Strom (2000) | ? | ? | ? | ? | – | + |
| Trompetter (2015) | + | ? | + | + | + | + |
| Williams (2010) | + | + | + | ? | + | + |
| Wilson (2015) | ? | ? | + | + | + | + |
| Yelvar (2017) | ? | + | ? | ? | + | + |

[a]The study satisfied the criteria.

[b]The study did not satisfy the criteria.

[c]Researchers were unable to determine if criteria were satisfied.

## Results of Individual Studies

The included studies indicate positive effects for interventions delivered via eHealth modalities in comparison with a control/enhanced control; 80% (24/30) of studies returned a reduction in pain interference, 69% (18/26) of studies returned a reduction in pain severity, 79% (19/24) of studies showed a decrease in psychological distress, and 67% (8/12) studies indicated an improvement in HRQoL.

Exploratory analyses were carried out on the primary outcome, pain interference, secondary outcomes, pain severity, psychological distress, and HRQoL. An NMA was conducted for the primary outcome, pain interference.

### *Exploratory Analysis*

Exploratory pairwise meta-analyses were conducted where possible (Table 4).

**Table 4.** Exploratory analyses.

| Comparison and outcome | Number of studies | Standardized mean difference | *P* value |
|---|---|---|---|
| **Internet versus control** | | | |
| Pain interference | 18 | 0.28 | <.001 |
| Pain severity | 15 | 0.2 | <.001 |
| Psychological distress | 16 | 0.35 | .001 |
| Health-related quality of life | 6 | 0.02 | .80 |
| **Internet versus enhanced control** | | | |
| Pain interference | 5 | 0.17 | .55 |
| Pain severity | 5 | 0.16 | .57 |
| Psychological distress | 4 | 0.14 | .33 |
| Health-related quality of life | 1 | 0.34 | .26 |

### Pain Interference

Data were extracted for pain interference, disability, functional interference, physical impairment, physical functioning, and headache disability, using a variety of measures: the Brief Pain Inventory (BPI), Visual Analogue Scale (VAS), Multidimensional Pain Inventory (MPI), Pain Disability Index, Survey of Pain Attitudes, Roland-Morris Disability Questionnaire, 36-Item Short-Form Health Survey (SF-36), Fibromyalgia Impact Questionnaire, Health Assessment Questionnaire–Disability Index, Profile of Chronic Pain-Screen, Headache Disability Index, Oswestry Disability Index, and the Migraine Disability Assessment. Pairwise meta-analyses indicate that internet-delivered interventions result in a small statistically significant reduction in pain interference when compared with a control group (*P*<.001).

### Pain Severity

Of the included studies, 26 included a measure of pain severity. Data were extracted for pain severity, pain intensity, average pain, typical pain, activity pain, and pain severity using the following measures: BPI, VAS, MPI, PCP-S, Pain Assessment Questionnaire, Brief Pain Questionnaire, Numeric Rating Scale, Visual Numeric Scale, McGill Pain Questionnaire, headache or pain diaries, and study-specific measures [6,67]. Internet-delivered studies returned a small statistically significant reduction in pain severity when compared with a control group (*P*<.001). A second NMA was conducted on the basis of the effectiveness of eHealth modalities in reducing pain severity; a network map and results are provided in Multimedia Appendix 3.

### Psychological Distress

Of the included studies, 24 used a measure of psychological distress. Data were extracted for depression, anxiety, mental health, and negative mood regulation using a variety of measures: the Hospital Anxiety and Depression Scale, Montgomery-Asberg Depression Rating Scale–Self Rated, Beck Depression Inventory, Negative Affect Scale, Depression Anxiety Stress Scales, Patient Health Questionnaire 9-Item, Personal Health Questionnaire Depression Scale, Short Form (SF) 8 Health Survey, SF-36, and Centre for Epidemiologic Studies Depression Scale. Internet-delivered interventions returned a statistically significant SMD of 0.35 when compared with a control (*P*=.001). Internet-delivered interventions did not return a statistically significant reduction in psychological distress when compared with an enhanced control (*P*=.33).

### Health-Related Quality of Life

Data were available on HRQoL for 12 studies. This was measured in a variety of ways, including the Quality of Life Interview, Patient Global Impression of Change, the Quality of Life Index, General Health Questionnaire 12-Item, 12-Item Short Form Survey, and SF-36. No statistically significant differences were found in HRQoL between the internet-delivered interventions and a control (*P*=.80).

## Presentation of Network Structure

The network map in Figure 2 demonstrates the available evidence for this reduction in pain interference network. For convenience, the circular nodes are eHealth modalities and the square nodes represent the control groups.

## Summary of Network Geometry

The available evidence was used to generate the network displayed in Figure 2. The number of studies behind each direct comparison is outlined in Table 5, which also includes the percentage of contribution that each comparison made to the entire network. As expected, the internet treatment versus control comparison contributes the highest percentage (17.67%) of evidence to the network. Some of the indirect comparisons required a long pathway to be generated (eg, comparing mobile apps with telephone-delivered interventions requires the direct evidence of the internet and control nodes). The comparisons based on longer paths were communicated with less precision [41].

**Figure 2.** Network map of electronic health modalities for chronic pain.

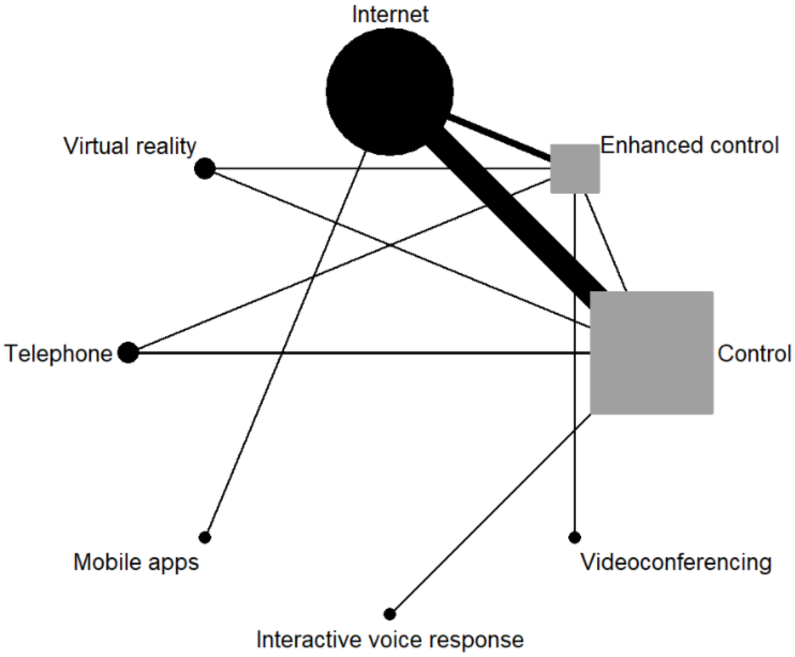



**Table 5.** Contribution to the pain interference network.

| Direct comparison | Studies, n | Contribution, % |
|---|---|---|
| Virtual reality versus control | 1 | 8.18 |
| Interactive voice response versus control | 1 | 12.10 |
| Internet versus control | 18 | 17.67 |
| Telephone versus control | 2 | 13.23 |
| Telephone versus enhanced control | 1 | 9.41 |
| Virtual reality versus enhanced control | 1 | 5.60 |
| Internet versus enhanced control | 5 | 3.11 |
| Videoconferencing versus enhanced control | 1 | 10.94 |
| Mobile apps versus internet | 1 | 10.74 |
| Enhanced control versus control | 1 | 9.01 |

## Synthesis of Results

### *Network Meta-Analysis (Pain Interference)*

A random-effects NMA based on the restricted maximum likelihood estimate was conducted to examine interventions delivered by eHealth modalities for the reduction of pain interference in chronic pain patients. The NMA suggests an SMD of 0.3, indicating a small difference between internet and the control (95% Credible Interval (CrI): 0.1 to 0.44) as expected by the exploratory analysis. In addition, an SMD of 0.28 was found between internet and the enhanced control (95% CrI: 0.002 to 0.55). The generated comparisons (Table 6) indicate that videoconferencing was significantly worse than all other modalities, bar interactive voice response.

The remaining comparisons had credible intervals containing 0, suggesting a high probability that the true comparison is not significant. Many of the credible intervals were very wide and expressed the uncertainty in the model's estimates. It must be stressed that many of these comparisons were based on a low sample size and do not suggest that significance cannot be achieved with a greater number of studies.

Table 7 outlines the rankings of the modalities and the probability that they deliver the most effective interventions. The mobile apps and virtual reality arms were given the best ranking, at second (95% CrI 1 to 7; 95% CrI 1 to 6). Slightly less uncertainty surrounds the ranking of the internet arm, with a median value of 3 and a credible interval from 1 to 5. The videoconferencing arm had a ranking of 8.

Table 7 indicates that there is a 43% chance that mobile app–delivered interventions are the most effective at reducing pain interference. The available evidence suggests that there is a 0% chance that videoconferencing delivers the most effective interventions.

**Table 6.** Results of network meta-analysis (NMA): electronic health (eHealth) modalities delivering interventions for reducing pain interference. Data in italics are statistically significant.

| Modality | Internet, SMD[a] (CrI[b]) | Virtual reality, SMD (CrI) | Telephone, SMD (CrI) | Mobile apps, SMD (CrI) | Interactive voice response, SMD (CrI) | Videoconferencing, SMD (CrI) | Enhanced control, SMD (CrI) |
|---|---|---|---|---|---|---|---|
| Virtual reality | −0.16 (−0.77 to 0.44) | —[c] | — | — | — | — | — |
| Telephone | −0.01 (−0.57 to 0.55) | 0.15 (−0.63 to 0.95) | — | — | — | — | — |
| Mobile apps | −0.21 (−0.95 to 0.54) | −0.04 (−1 to 0.92) | −0.19 (−1.12 to 0.74) | — | — | — | — |
| Interactive voice response | 0.39 (−0.44 to 0.12) | 0.55 (−0.46 to 1.57) | 0.4 (−0.59 to 1.39) | 0.6 (−0.52 to 1.72) | — | — | — |
| Videoconferencing | *1.59 (0.82 to 2.36)* | *1.75 (0.81 to 2.69)* | *1.6 (0.67 to 2.54)* | *1.8 (0.72 to 2.87)* | *1.2 (0.06 to 2.33)* | — | — |
| Enhanced control | *0.28 (0.002 to 0.55)* | 0.44 (−0.16 to 1.03) | 0.29 (−0.29 to 0.87) | 0.49 (−0.31 to 1.27) | −0.11 (−0.99 to 0.76) | *-1.32 (-2.1 to −0.53)* | — |
| Control | *0.3 (0.1 to 0.44)* | 0.44 (−0.15 to 1.04) | 0.29 (−0.26 to 0.83) | 0.48 (−0.28 to 1.24) | −0.11 (−0.93 to 0.7) | *− 1.31 (−2 to −0.59)* | −0.004 (−0.31 to 0.31) |

[a]SMD: standardized mean difference.

[b]CrI: credible interval.

[c]Not applicable.

**Table 7.** Ranked effectiveness of modalities.

| Modality | Median ranking[a] (credible interval) | Probability[b] (SD) |
|---|---|---|
| Internet | 3 (1 to 5) | .04 (0.19) |
| Virtual reality | 2 (1 to 6) | .34 (0.47) |
| Telephone | 3 (1 to 7) | .15 (0.36) |
| Mobile apps | 2 (1 to 7) | .43 (0.50) |
| Interactive voice response | 7 (1 to 7) | .05 (0.21) |
| Videoconferencing | 8 (8 to 8) | .000008 (0.003) |
| Enhanced control | 6 (3 to 7) | .0007 (0.03) |
| Control | 6 (4 to 7) | .00002 (0.004) |

[a]Treatments ranked in order of comparative effectiveness.

[b]Probability of each treatment being the best (ie, most effective).

## Exploration for Inconsistency

Given that all of the studies included in this review were randomized, the assumption of transitivity is fulfilled. A test of loop inconsistency and a Lu-Ades test of design inconsistency revealed no evidence of inconsistency ($P$=.85 and $P$=.67, respectively). Node splitting returned no evidence of inconsistency when assessing differences between direct and indirect effects.

## Risk of Bias Across Studies

The funnel plot (see Multimedia Appendix 4) showed no indication of publication bias with the majority of studies falling within the bands. Most studies were clustered around the zero line and have relatively large SEs.

## Results of Additional Analyses

Sensitivity analyses were carried out to assess the fit of the model. A variety of different initial values were tested; the model was run with an extended burn-in of 200,000 iterations. Both a gamma and half normal prior were used to ensure that the normal prior was uninformative, and 500,000 and 700,000 iterations were run to ensure that 600,000 were adequate. In addition, the model was run with 2 chains, and history plots showed tight iterations, indicating no evidence of nonconvergence.

Additional covariates were added to the model to explore heterogeneity. The initial NMA model returned a DIC of 12.47. When the covariates were added to the model, the DIC did not significantly reduce and they were considered not to have added enough to the model to warrant inclusion (age [DIC=13.72], gender [DIC=12.79], length of intervention [DIC=12.84], attrition [DIC=13.15], measure [DIC=13.20], contact [DIC=13.03], analysis [DIC=12.99], and condition [DIC=12.99]).

## Discussion

### Principal Findings

The random-effects NMA returned pairwise comparisons between each of the eHealth modalities. The majority of these comparisons were not statistically significant; however, the network indicates that all eHealth modalities were significantly better than videoconferencing. On the basis of the currently available evidence, the network also promotes the use of internet to deliver interventions. This study created a ranked list of eHealth modalities used for chronic pain by conducting a systematic review with an NMA. Study findings tentatively indicated that mobile apps and virtual reality were the most effective eHealth modalities for delivering interventions for reducing pain interference. More specifically, the joint highest ranked modalities overall, according to NMA analyses, were mobile apps, with a 43% chance that this modality delivered the most effective intervention for reducing pain interference. Following this, virtual reality had a 34% chance and telephone had a 15% chance of being the most effective delivery method. Internet-delivered interventions have a 4% chance of being the most effective at reducing pain; however, there was more certainty regarding their positioning and effectiveness as they contributed the most papers [27] to the network (comparisons including internet contributed a total of 20.78% to the network). Although the analyses revealed important insights for the potential rank order of eHealth modalities for chronic pain interventions, only tentative conclusions regarding the most effective treatment types can be drawn, as there are limitations with this review.

### Strengths and Limitations

One limitation with this review is the disproportionate representation of different eHealth modalities included within the network. For example, of the 30 papers included in the analysis, internet was represented in 23 papers and telephone was represented in 2 papers, whereas, mobile apps, interactive voice response, and videoconferencing were each represented in only 1 paper. As a result, although we can be confident of the ranking of internet relative to the other modalities in the network, we cannot be confident of the rankings of the other modalities relative to internet. To explain further, if for example, an additional internet paper was added to this network, the modality rankings would not be anticipated to change, but if a new study based on another modality was added, then there is a chance that the modality rankings would change. However, this review is bound by the available evidence, and the current synthesis provides the first steps toward ranking which eHealth methodologies are more efficacious in the context of chronic pain.

It must also be noted that a contributing factor to the limited number of included papers and, therefore, eHealth modality types may have been the restrictive inclusion and exclusion criteria used in this study. For example, 51 studies were excluded for not being RCTs and 13 studies were excluded for not having 20 participants per arm for each time point. Therefore, had the eligibility criteria been more relaxed, arguably, more studies would have been included, allowing a larger network to be produced. However, the inclusion and exclusion criteria employed in this review followed on from a previous review in the area and the exacting criteria ensured that the included papers were of high quality and had low risk of bias [49].

Finally, because of the heterogeneous nature of intervention content, it may be contentious whether the current approach was optimal to identify the effectiveness of eHealth modalities relative to one another. For example, if each study in this review administered cognitive behavioral therapy (CBT) across each modality type, through accounting for differences in extracted variables (eg, age and gender), it could be reasonably assumed that any notable differences detected were because of the effect of the modality and not the intervention content (ie, CBT). However, although this was not the case in this network, it may also be debated that the aforementioned scenario would actually yield which eHealth modality is best for a particular treatment type (eg, CBT) and not which eHealth modality is most efficacious in the context of chronic pain. In any case, the scientific and clinical purposes of this review were to identify which eHealth modality, on the basis of the available evidence, delivers the most efficacious intervention for people living with chronic pain and not which intervention type (eg, CBT) works best with which eHealth modality.

Although there are certain limitations with this review, the findings provide support for previous research, yield tentative conclusions regarding the ranked efficacy of eHealth modalities in the context of chronic pain, and offer insight into further areas for investigation. Similar to previous research [49], the results from the exploratory meta-analysis highlight that internet-delivered interventions can reduce pain interference for people living with chronic pain. Interestingly, with regard to the results from the NMA, the 2 modalities found, albeit tentatively, to be most efficacious (virtual reality and mobile apps) are relatively new eHealth modalities compared with others in the review. The reason is not clear, but perhaps these modalities offer more immersive and convenient intervention pathways that appeal to participants.

Empowering individuals to take an active role in their own health care has been identified as a crucial factor for improving the quality of care and reducing health care costs [73-76]. This is particularly important for people with long-term health issues that require prolonged lifestyle modifications and adjustments [77-79]. Self-management of chronic/long-term illnesses through education and supportive interventions can not only decrease utilization of health care services but may also lead to improvements in clinical outcomes and overall quality of life [73]. Increasing patient engagement in health care interventions has thus become a priority for health care organizations, researchers, and policy makers. eHealth modalities offer tremendous potential to engage patients as they are flexible and can be tailored to individual patient's needs, preferences, and circumstances [78]. However, as these technologies require actions that must be initiated and sustained by the individual, it is vital that these interventions are designed in an easily accessible and engaging manner. Perhaps as the research findings of this study tentatively show, interventions delivered via virtual reality and mobile apps could yield promising results in this area by virtue of their immersive and accessible design. In particular, research should focus on conducting interventions with mobile apps for chronic pain. With 93% of Irish consumers having access to a mobile phone [80], and a myriad of mobile apps targeting people with chronic pain (a recent review found 373 mobile apps for older adults with arthritic pain alone [81]), it is concerning that only 1 study included in this review delivered an intervention via a mobile app.

### Conclusions

In the wider context of eHealth, there are 2 areas for future work. The first would be to replicate the synthesis of this review with different chronic conditions. The second area for future work would be to create a core outcome set for eHealth interventions, a standardized set of eHealth intervention engagement outcomes measuring, for example, fidelity, participant engagement, and user experience. Often, a treatment can have an effect in person, but this effect may not transfer to an eHealth intervention. In such instances, it is quite possible that the eHealth execution and delivery was unsatisfactory and *not* that the intervention content cannot be adapted to an eHealth version. A core eHealth outcome set would assist in negating such issues.

In conclusion, from both a clinical and scientific perspective, previous research has outlined a need to compare eHealth modalities in the context of chronic pain. This research is the first to use a novel statistical method, namely, NMA, to quantitatively compare eHealth modalities in this context. Similar to previous research, the results suggest that internet interventions can improve pain interference, whereas more novel modalities (ie, mobile apps and virtual reality) are most likely to be effective, but more research on chronic pain eHealth is needed. Among many areas for future research, additional research examining underutilized eHealth modalities is recommended, and a core outcome set with regard to measuring engagement within eHealth interventions in general is paramount.

## Acknowledgments

This work is supported by the Irish Health Research Board Research Leaders Award (grant reference: BEM, RLA/2013).

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Detailed definitions/explanation of each electronic health modality.

[PDF File (Adobe PDF File), 75KB - jmir_v21i7e11086_app1.pdf ]

## Multimedia Appendix 2

Preferred Reporting Items for Systematic Reviews and Meta-analyses network meta-analysis checklist of items to include when reporting a systematic review involving a network meta-analysis.

[PDF File (Adobe PDF File), 140KB - jmir_v21i7e11086_app2.pdf ]

## Multimedia Appendix 3

Network meta-analysis of electronic health modalities used to deliver interventions for the reduction of pain severity in a chronic pain population.

[PDF File (Adobe PDF File), 60KB - jmir_v21i7e11086_app3.pdf ]

## Multimedia Appendix 4

Risk of bias across studies.

[PDF File (Adobe PDF File), 39KB - jmir_v21i7e11086_app4.pdf ]

## References

1. Slattery BW, Haugh S, Francis K, O'Connor L, Barrett K, Dwyer CP, et al. Protocol for a systematic review with network meta-analysis of the modalities used to deliver eHealth interventions for chronic pain. Syst Rev 2017 Dec 03;6(1):45 [FREE Full text] [doi: 10.1186/s13643-017-0414-x] [Medline: 28253909]
2. Showell C, Nøhr C. How should we define eHealth, and does the definition matter? Stud Health Technol Inform 2012;180:881-884. [Medline: 22874319]

3. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
4. Rosser BA, Vowles KE, Keogh E, Eccleston C, Mountain GA. Technologically-assisted behaviour change: a systematic review of studies of novel technologies for the management of chronic illness. J Telemed Telecare 2009;15(7):327-338. [doi: 10.1258/jtt.2009.090116] [Medline: 19815901]
5. Krein SL, Kadri R, Hughes M, Kerr EA, Piette JD, Holleman R, et al. Pedometer-based internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. J Med Internet Res 2013 Aug 19;15(8):e181 [FREE Full text] [doi: 10.2196/jmir.2605] [Medline: 23969029]
6. Leveille SG, Huang A, Tsai SB, Allen M, Weingart SN, Iezzoni LI. Health coaching via an internet portal for primary care patients with chronic conditions: a randomized controlled trial. Med Care 2009 Jan;47(1):41-47. [doi: 10.1097/MLR.0b013e3181844dd0] [Medline: 19106729]
7. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
8. Ström L, Pettersson R, Andersson G. A controlled trial of self-help treatment of recurrent headache conducted via the internet. J Consult Clin Psychol 2000 Aug;68(4):722-727. [Medline: 10965647]
9. Wilson M, Roll JM, Corbett C, Barbosa-Leiker C. Empowering patients with persistent pain using an internet-based self-management program. Pain Manag Nurs 2015 Aug;16(4):503-514. [doi: 10.1016/j.pmn.2014.09.009] [Medline: 26088940]
10. Andersson G, Lundström P, Ström L. Internet-based treatment of headache: does telephone contact add anything? Headache 2003 Apr;43(4):353-361. [doi: 10.1046/j.1526-4610.2003.03070.x] [Medline: 12656706]
11. Heapy A, Sellinger J, Higgins D, Chatkoff D, Bennett TC, Kerns RD. Using interactive voice response to measure pain and quality of life. Pain Med 2007 Oct 01;8(suppl 3):S145-S154 [FREE Full text] [doi: 10.1111/j.1526-4637.2007.00378.x]
12. Naylor MR, Keefe FJ, Brigidi B, Naud S, Helzer JE. Therapeutic interactive voice response for chronic pain reduction and relapse prevention. Pain 2008 Feb;134(3):335-345 [FREE Full text] [doi: 10.1016/j.pain.2007.11.001] [Medline: 18178011]
13. Botella C, Garcia-Palacios A, Vizcaíno Y, Herrero R, Baños RM, Belmonte MA. Virtual reality in the treatment of fibromyalgia: a pilot study. Cyberpsychol Behav Soc Netw 2013 Mar;16(3):215-223. [doi: 10.1089/cyber.2012.1572] [Medline: 23496678]
14. Garcia-Palacios A, Herrero R, Vizcaíno Y, Belmonte MA, Castilla D, Molinari G, et al. Integrating virtual reality with activity management for the treatment of fibromyalgia: acceptability and preliminary efficacy. Clin J Pain 2015 Jun;31(6):564-572. [doi: 10.1097/AJP.0000000000000196] [Medline: 25551475]
15. Herbert MS, Afari N, Liu L, Heppner P, Rutledge T, Williams K, et al. Telehealth versus in-person acceptance and commitment therapy for chronic pain: a randomized noninferiority trial. J Pain 2017 Feb;18(2):200-211. [doi: 10.1016/j.jpain.2016.10.014] [Medline: 27838498]
16. 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]
17. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. J Med Internet Res 2015;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]
18. Reed JL, Prince SA, Cole CA, Nerenberg KA, Hiremath S, Tulloch HE, et al. E-health physical activity interventions and moderate-to-vigorous intensity physical activity levels among working-age women: a systematic review protocol. Syst Rev 2015 Jan 14;4:3 [FREE Full text] [doi: 10.1186/2046-4053-4-3] [Medline: 25589330]
19. Capurro D, Ganzinger M, Perez-Lu J, Knaup P. Effectiveness of eHealth interventions and information needs in palliative care: a systematic literature review. J Med Internet Res 2014;16(3):e72 [FREE Full text] [doi: 10.2196/jmir.2812] [Medline: 24610324]
20. Muellmann S, Forberger S, Möllers T, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review protocol. Syst Rev 2016 Mar 16;5:47 [FREE Full text] [doi: 10.1186/s13643-016-0223-7] [Medline: 26984189]
21. Heapy AA, Higgins DM, Cervone D, Wandner L, Fenton BT, Kerns RD. A systematic review of technology-assisted self-management interventions for chronic pain. Clin J Pain 2015 Jun;31(6):470-492. [doi: 10.1097/AJP.0000000000000185] [Medline: 25411862]
22. Liaw S, Humphreys JS. Rural eHealth paradox: it's not just geography!. Aust J Rural Health 2006 Jun;14(3):95-98. [doi: 10.1111/j.1440-1584.2006.00786.x] [Medline: 16706876]
23. Stroetmann K, Jones T, Dobrev A, Stroetmann V. European Commission. Luxembourg: Office for Official Publications of the European Communities; 2006. eHealth is Worth it. The economic benefits of implemented eHealth solutions at ten European sites URL: https://tinyurl.com/yxkc7goa [accessed 2019-03-25] [WebCite Cache ID 778bMtPaW]
24. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]

25. Talboom-Kamp EP, Verdijk NA, Harmans LM, Numans ME, Chavannes NH. An eHealth platform to manage chronic disease in primary care: an innovative approach. Interact J Med Res 2016 Feb 09;5(1):e5 [FREE Full text] [doi: 10.2196/ijmr.4217] [Medline: 26860333]
26. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
27. Kelley H, Chiasson M, Downey A, Pacaud D. The clinical impact of eHealth on the self-management of diabetes: a double adoption perspective. J Assoc Inf Syst 2011 Mar;12(3):208-234. [doi: 10.17705/1jais.00263]
28. Nes AA, van Dulmen S, Eide E, Finset A, Kristjánsdóttir OB, Steen IS, et al. The development and feasibility of a web-based intervention with diaries and situational feedback via smartphone to support self-management in patients with diabetes type 2. Diabetes Res Clin Pract 2012 Sep;97(3):385-393. [doi: 10.1016/j.diabres.2012.04.019] [Medline: 22578890]
29. Villani D, Cognetta C, Toniolo D, Scanzi F, Riva G. Engaging elderly breast cancer patients: the potential of eHealth interventions. Front Psychol 2016;7:1825 [FREE Full text] [doi: 10.3389/fpsyg.2016.01825] [Medline: 27899909]
30. Green BB, Anderson ML, Ralston JD, Catz S, Fishman PA, Cook AJ. Patient ability and willingness to participate in a web-based intervention to improve hypertension control. J Med Internet Res 2011 Jan 20;13(1):e1 [FREE Full text] [doi: 10.2196/jmir.1625] [Medline: 21371993]
31. Bogosian A, Chadwick P, Windgassen S, Norton S, McCrone P, Mosweu I, et al. Distress improves after mindfulness training for progressive MS: a pilot randomised trial. Mult Scler 2015 Aug;21(9):1184-1194. [doi: 10.1177/1352458515576261] [Medline: 25767124]
32. Buhrman M, Fredriksson A, Edström G, Shafiei D, Tärnqvist C, Ljótsson B, et al. Guided internet-delivered cognitive behavioural therapy for chronic pain patients who have residual symptoms after rehabilitation treatment: randomized controlled trial. Eur J Pain 2013 May;17(5):753-765. [doi: 10.1002/j.1532-2149.2012.00244.x] [Medline: 23139021]
33. Buhrman M, Syk M, Burvall O, Hartig T, Gordh T, Andersson G. Individualized guided internet-delivered cognitive-behavior therapy for chronic pain patients with comorbid depression and anxiety: a randomized controlled trial. Clin J Pain 2015 Jun;31(6):504-516. [doi: 10.1097/AJP.0000000000000176] [Medline: 25380222]
34. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
35. Kroenke K, Krebs EE, Wu J, Yu Z, Chumbler NR, Bair MJ. Telecare collaborative management of chronic pain in primary care: a randomized clinical trial. J Am Med Assoc 2014 Jul 16;312(3):240-248. [doi: 10.1001/jama.2014.7689] [Medline: 25027139]
36. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KMG. Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
37. Williams DA, Kuper D, Segar M, Mohan N, Sheth M, Clauw DJ. Internet-enhanced management of fibromyalgia: a randomized controlled trial. Pain 2010 Dec;151(3):694-702 [FREE Full text] [doi: 10.1016/j.pain.2010.08.034] [Medline: 20855168]
38. Stinson JN, Lalloo C, Harris L, Isaac L, Campbell F, Brown S, et al. iCanCope with Pain™: user-centred design of a web- and mobile-based self-management program for youth with chronic pain based on identified health care needs. Pain Res Manag 2014;19(5):257-265 [FREE Full text] [Medline: 25000507]
39. Raftery MN, Sarma K, Murphy AW, De la Harpe D, Normand C, McGuire BE. Chronic pain in the Republic of Ireland--community prevalence, psychosocial profile and predictors of pain-related disability: results from the Prevalence, Impact and Cost of Chronic Pain (PRIME) study, part 1. Pain 2011 May;152(5):1096-1103. [doi: 10.1016/j.pain.2011.01.019] [Medline: 21450402]
40. McGuire BE, Henderson EM, McGrath PJ. Translating e-pain research into patient care. Pain 2017 Dec;158(2):190-193. [doi: 10.1097/j.pain.0000000000000686] [Medline: 27525833]
41. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health 2014 Mar;17(2):157-173 [FREE Full text] [doi: 10.1016/j.jval.2014.01.004] [Medline: 24636374]
42. Naci H, Ioannidis JP. Comparative effectiveness of exercise and drug interventions on mortality outcomes: metaepidemiological study. Br Med J 2013 Oct 01;347:f5577 [FREE Full text] [doi: 10.1136/bmj.f5577] [Medline: 24473061]
43. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011 Jun;14(4):417-428 [FREE Full text] [doi: 10.1016/j.jval.2011.04.002] [Medline: 21669366]
44. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015 Jun 02;162(11):777-784. [doi: 10.7326/M14-2385] [Medline: 26030634]

45. Slattery B, Haugh S, Francis K, O?Connor L, Barrett K, McGuire B. National Institute of Health Research. York, United Kingdom: PROSPERO International prospective register of systematic reviews; 2016. An evaluation of the modalities used to deliver eHealth interventions for chronic pain: a systematic review with network meta-analysis URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=35595 [accessed 2019-03-25] [WebCite Cache ID 778blaupq]
46. Berman RL, Iris MA, Bode R, Drengenberg C. The effectiveness of an online mind-body intervention for older adults with chronic pain. J Pain 2009 Jan;10(1):68-79. [doi: 10.1016/j.jpain.2008.07.006] [Medline: 18774342]
47. Watts TL. Probing site configuration in patients with untreated periodontitis. A study of horizontal positional error. J Clin Periodontol 1989 Sep;16(8):529-533. [doi: 10.1111/j.1600-051X.1989.tb02331.x] [Medline: 2778087]
48. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003 Dec;106(3):337-345. [Medline: 14659516]
49. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (Internet-delivered) for the management of chronic pain in adults. Cochrane Database Syst Rev 2014;2:CD010152. [doi: 10.1002/14651858.CD010152.pub2] [Medline: 24574082]
50. Biomed Central. ISRCTN URL: http://www.isrctn.com/search?q= [accessed 2018-05-18] [WebCite Cache ID 6zVJPK6LC]
51. The National Library of Medicine. ClinicalTrials URL: http://www.clinicaltrials.gov [accessed 2018-05-18] [WebCite Cache ID 6zVIjStA6]
52. World Health Organization. 2018. International Clinical Trials Registry Platform URL: http://apps.who.int/trialsearch/ [accessed 2018-05-18] [WebCite Cache ID 6zVJ9ewTS]
53. White IR, Barrett JK, Jackson D, Higgins JP. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Methods 2012 Jun;3(2):111-125 [FREE Full text] [doi: 10.1002/jrsm.1045] [Medline: 26062085]
54. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods 2012 Jun;3(2):98-110 [FREE Full text] [doi: 10.1002/jrsm.1044] [Medline: 26062084]
55. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making 2013 Dec;33(5):641-656 [FREE Full text] [doi: 10.1177/0272989X12455847] [Medline: 23804508]
56. de Boer M, Versteegen GJ, Vermeulen KM, Sanderman R, Struys MM. A randomized controlled trial of an Internet-based cognitive-behavioural intervention for non-specific chronic pain: an effectiveness and cost-effectiveness study. Eur J Pain 2014 Nov;18(10):1440-1451. [doi: 10.1002/ejp.509] [Medline: 24777973]
57. Bromberg J, Wood ME, Black RA, Surette DA, Zacharoff KL, Chiauzzi EJ. A randomized trial of a web-based intervention to improve migraine self-management and coping. Headache 2012 Feb;52(2):244-261 [FREE Full text] [Medline: 22413151]
58. Buhrman M, Fältenhag S, Ström L, Andersson G. Controlled trial of internet-based treatment with telephone support for chronic back pain. Pain 2004 Oct;111(3):368-377. [doi: 10.1016/j.pain.2004.07.021] [Medline: 15363881]
59. Buhrman M, Nilsson-Ihrfeldt E, Jannert M, Ström L, Andersson G. Guided internet-based cognitive behavioural treatment for chronic back pain reduces pain catastrophizing: a randomized controlled trial. J Rehabil Med 2011 May;43(6):500-505 [FREE Full text] [doi: 10.2340/16501977-0805] [Medline: 21533329]
60. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
61. Dear BF, Titov N, Perry KN, Johnston L, Wootton BM, Terides MD, et al. The Pain Course: a randomised controlled trial of a clinician-guided Internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. Pain 2013 Jun;154(6):942-950. [doi: 10.1016/j.pain.2013.03.005] [Medline: 23688830]
62. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The Pain Course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. Pain 2015 Oct;156(10):1920-1935 [FREE Full text] [doi: 10.1097/j.pain.0000000000000251] [Medline: 26039902]
63. Dear BF, Gandy M, Karin E, Ricciardi T, Fogliati VJ, McDonald S, et al. The pain course: a randomised controlled trial comparing a remote-delivered chronic pain management program when provided in online and workbook formats. Pain 2017 Jul;158(7):1289-1301. [doi: 10.1097/j.pain.0000000000000916] [Medline: 28394850]
64. Devineni T, Blanchard EB. A randomized controlled trial of an internet-based treatment for chronic headache. Behav Res Ther 2005 Mar;43(3):277-292. [doi: 10.1016/j.brat.2004.01.008] [Medline: 15680926]
65. Kleiboer A, Sorbi M, van Silfhout M, Kooistra L, Passchier J. Short-term effectiveness of an online behavioral training in migraine self-management: a randomized controlled trial. Behav Res Ther 2014 Oct;61:61-69. [doi: 10.1016/j.brat.2014.07.009] [Medline: 25133856]
66. Müller R, Gertz KJ, Molton IR, Terrill AL, Bombardier CH, Ehde DM, et al. fEffects of a tailored positive psychology intervention on well-being and pain in individuals with chronic pain and a physical disability: a feasibility trial. Clin J Pain 2016 Jan;32(1):32-44. [doi: 10.1097/AJP.0000000000000225] [Medline: 25724020]
67. Peters ML, Smeets E, Feijge M, van Breukelen G, Andersson G, Buhrman M, et al. Happy Despite Pain: a randomized controlled trial of an 8-week internet-delivered positive psychology intervention for enhancing well-being in patients with

chronic pain. Clin J Pain 2017 Nov;33(11):962-975 [FREE Full text] [doi: 10.1097/AJP.0000000000000494] [Medline: 28379873]
68. Ruehlman LS, Karoly P, Enders C. A randomized controlled evaluation of an online chronic pain self management program. Pain 2012 Feb;153(2):319-330 [FREE Full text] [doi: 10.1016/j.pain.2011.10.025] [Medline: 22133450]
69. Lin J, Paganini S, Sander L, Lüking M, Ebert DD, Buhrman M, et al. An internet-based intervention for chronic pain. Dtsch Arztebl Int 2017 Oct 13;114(41):681-688 [FREE Full text] [doi: 10.3238/arztebl.2017.0681] [Medline: 29082858]
70. McBeth J, Prescott G, Scotland G, Lovell K, Keeley P, Hannaford P, et al. Cognitive behavior therapy, exercise, or both for treating chronic widespread pain. Arch Intern Med 2012 Jan 9;172(1):48-57. [doi: 10.1001/archinternmed.2011.555] [Medline: 22082706]
71. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
72. Yilmaz YG, Çırak Y, Dalkılınç M, Parlak DY, Guner Z, Boydak A. Is physiotherapy integrated virtual walking effective on pain, function, and kinesiophobia in patients with non-specific low-back pain? Randomised controlled trial. Eur Spine J 2017 Dec;26(2):538-545. [doi: 10.1007/s00586-016-4892-7] [Medline: 27981455]
73. Vahdat S, Hamzehgardeshi L, Hessam S, Hamzehgardeshi Z. Patient involvement in health care decision making: a review. Iran Red Crescent Med J 2014 Jan;16(1):e12454 [FREE Full text] [doi: 10.5812/ircmj.12454] [Medline: 24719703]
74. Clancy CM. Patient engagement in health care. Health Serv Res 2011 Apr;46(2):389-393 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01254.x] [Medline: 21371026]
75. Gruman J, Rovner MH, French ME, Jeffress D, Sofaer S, Shaller D, et al. From patient education to patient engagement: implications for the field of patient education. Patient Educ Couns 2010 Mar;78(3):350-356. [doi: 10.1016/j.pec.2010.02.002] [Medline: 20202780]
76. 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]
77. 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]
78. Fleming MD, Shim JK, Yen IH, Thompson-Lastad A, Rubin S, Van Natta M, et al. Patient engagement at the margins: health care providers' assessments of engagement and the structural determinants of health in the safety-net. Soc Sci Med 2017 Dec;183:11-18 [FREE Full text] [doi: 10.1016/j.socscimed.2017.04.028] [Medline: 28445806]
79. Mullins CD, Abdulhalim AM, Lavallee DC. Continuous patient engagement in comparative effectiveness research. J Am Med Assoc 2012 Apr 18;307(15):1587-1588. [doi: 10.1001/jama.2012.442] [Medline: 22511684]
80. Deloitte. Dublin: Deloitte Ireland LLP; 2018. Mobile Consumer Survey 2018: The Irish Cut URL: https://tinyurl.com/yx94os4k [accessed 2019-03-25] [WebCite Cache ID 778ebzieG]
81. Bhattarai P, Newton-John TR, Phillips JL. Quality and Usability of Arthritic Pain Self-Management Apps for Older Adults: A Systematic Review. Pain Med 2018 Mar 01;19(3):471-484. [doi: 10.1093/pm/pnx090] [Medline: 28541464]

## Abbreviations

**BPI:** Brief Pain Inventory
**CBT:** cognitive behavioral therapy
**CENTRAL:** Cochrane Central Register of Controlled Trials
**CrI:** Credible Interval
**DIC:** deviance information criterion
**eHealth:** electronic health
**EMBASE:** Excerpta Medica dataBASE
**HRQoL:** health-related quality of life
**MEDLINE:** Medical Literature Analysis and Retrieval System Online
**MPI:** Multidimensional Pain Inventory
**NMA:** network meta-analysis
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RCT:** randomized controlled trial
**SMD:** standardized mean difference
**VAS:** Visual Analogue Scale

*Edited by G Eysenbach; submitted 18.05.18; peer-reviewed by B Thompson, H Devan, J Priebe, H Slater, R Van der Vaart; comments to author 30.08.18; revised version received 20.12.18; accepted 06.01.19; published 17.07.19.*

*Please cite as:*
*Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, Egan J, McGuire BE*
*An Evaluation of the Effectiveness of the Modalities Used to Deliver Electronic Health Interventions for Chronic Pain: Systematic Review With Network Meta-Analysis*
*J Med Internet Res 2019;21(7):e11086*
*URL: https://www.jmir.org/2019/7/e11086/*
*doi:10.2196/11086*
*PMID:31317869*

Original Paper

# Association Between Physical Activity Intervention Website Use and Physical Activity Levels Among Spanish-Speaking Latinas: Randomized Controlled Trial

Sarah E Linke[1], MPH, PhD; Shira I Dunsiger[2], PhD; Kim M Gans[3,4], PhD; Sheri J Hartman[1], PhD; Dori Pekmezi[5], PhD; Britta A Larsen[1], PhD; Andrea S Mendoza-Vasconez[1], MMPH; Bess H Marcus[2,4], PhD

[1]Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, United States

[2]Centers for Behavioral and Preventive Medicine, Department of Psychiatry and Human Behavior, Miriam Hospital and Warren Alpert Medical School, Brown University, Providence, RI, United States

[3]Department of Human Development and Family Studies, Institute for Collaboration on Health, Intervention, and Policy, University of Connecticut, Storrs, CT, United States

[4]Department of Behavioral and Social Sciences and the Institute for Community Health Promotion, School of Public Health, Brown University, Providence, RI, United States

[5]Department of Health Behavior, School of Public Health, University of Alabama, Birmingham, AL, United States

**Corresponding Author:**
Sarah E Linke, MPH, PhD
Department of Family Medicine and Public Health
University of California, San Diego
9500 Gilman Drive
Box 0725
La Jolla, CA,
United States
Phone: 1 858 822 1082
Email: slinke@ucsd.edu

## Abstract

**Background:** The internet's low cost and potential for high reach makes Web-based channels prime for delivering evidence-based physical activity (PA) interventions. Despite the well-studied success of internet-based PA interventions in primarily non-Hispanic white populations, evidence on Spanish-speaking Latinas' use of such interventions is lacking. The recent rise in technology use among Latinas in the United States, a population at heightened risk for low PA levels and related conditions, suggests that they may benefit from Web-based PA interventions tailored to their cultural and language preferences.

**Objective:** The goal of the research was to examine participant engagement with various features of an internet-based PA intervention for Latinas and explore how use of these features was differentially associated with adoption and maintenance of PA behavior change.

**Method:** Pasos Hacia la Salud tested a Spanish-language, culturally adapted, individually tailored, internet-based PA intervention versus a Spanish language, internet-based, Wellness Contact Control condition for underactive Latinas (N=205, mean age 39.2 [SD 10.5] years, 84% Mexican American). These analyses examined engagement with the website and explored how use was associated with adoption and maintenance of moderate to vigorous physical activity (MVPA) behavior.

**Results:** Overall, participants logged on to the website an average of 22 times (SD 28) over 12 months, with intervention participants logging on significantly more than controls (29 vs 14.7, $P<.001$). On average, participants spent more time on the website at months 1, 4, and 6 compared to all other months, with maximum use at month 4. Both log-ins and time spent on the website were significantly related to intervention success (achieving higher mean minutes of MVPA per week at follow-up: $b=.48$, SE 0.20, $P=.02$ for objectively measured MVPA and $b=.74$, SE 0.34, $P=.03$ for self-reported MVPA at 12 months, controlling for baseline). Furthermore, those meeting guidelines by the Centers for Disease Control and Prevention for PA at 12 months (≥150 minutes per week of MVPA) logged on significantly more than those not meeting guidelines (35 vs 20 over 12 months, $P=.002$). Among participants in the intervention arm, goal-setting features, personal PA reports, and PA tips were the most used portions of the website. Higher use of these features was associated with greater success in the program (significantly more minutes of self-reported MVPA at 12 months controlling for baseline). Specifically, one additional use of these features per month over 12 months translated into an additional 34 minutes per week of MVPA (goals feature), 12 minutes per week (PA tips), and 42 minutes per week (PA reports).

**Conclusions:** These results demonstrate that greater use of a tailored, Web-based PA intervention, particularly certain features on the site, was significantly related to increased PA levels in Latinas.

**Trial Registration:** ClinicalTrials.gov NCT01834287; https://clinicaltrials.gov/ct2/show/NCT01834287

***(J Med Internet Res 2019;21(7):e13063)*** doi:10.2196/13063

**KEYWORDS**

physical activity; Latinas; internet; treatment engagement

## Introduction

Robust evidence supports physical activity (PA) in the prevention and management of numerous chronic health conditions, including obesity, hypertension, type 2 diabetes, cardiovascular diseases, and many types of cancer [1]. Despite these well-established benefits of regular PA, most Americans are insufficiently active [2]. Interventions designed to promote PA have had mixed success, and even those that have resulted in overall increases in PA have shown a range of adherence to behavior change among individuals [3]. A better understanding of the intervention features that successfully promote PA can inform the development of more efficient and broadly effective interventions.

Treatment engagement is a consistent predictor of success in health behavior interventions [4-6]. It is analogous to taking medication as prescribed; if the proper dose is not taken, the medication will be less effective. One of the greatest challenges facing behavioral scientists, therefore, is designing interventions to optimize treatment engagement and adherence and, in effect, maximize efficacy [7]. Delivering intervention materials and information in convenient and efficient ways may help to increase their uptake.

Use of technology, such as the Web, could be a promising solution for increasing engagement. Internet-based interventions offer numerous benefits for researchers and participants such as convenience, portability, adaptability, cost effectiveness, and reach. Internet-based interventions can incorporate multiple behavioral adherence strategies such as self-monitoring, social support, and goal-setting in a more accessible manner than in-person or print-based health behavior change interventions, and the interactive features of websites could promote participant engagement [4]. Additionally, internet-based interventions allow for objective measures of participant engagement by tracking variables such as number of log-ins, page clicks, and time spent on various components of the intervention. As internet-based interventions have achieved mixed success (ie, some are successful and others are not) [8], a more thorough examination of these objective measures of engagement could be especially useful in illuminating which intervention features promote not only engagement in the intervention but also successful behavior change.

In addition to promoting and tracking participant engagement, internet-based interventions could be an especially effective delivery channel for racial and ethnic minorities. Internet use in Latinos has grown markedly in recent years [9]. In 2015, 84% of Latinos were online with the fastest growing use among immigrant Hispanics and those who are Spanish dominant [10]. The Latino population also reports lower levels of PA than non-Latino whites [11]. Latina women report less activity than non-Latino white women and Latino men, and Latinas also experience higher rates of obesity, diabetes, and other chronic conditions related to inactive lifestyle [11]. Given the rise in internet use among Latinas and the potential of internet-based interventions to incorporate features associated with successful behavior change, internet-based interventions may be particularly appropriate for promoting PA in Latinas.

The purpose of this paper was to examine participant engagement with various features of an internet-based PA intervention for Latinas and explore how use of these features was differentially associated with adoption and maintenance of PA behavior change.

## Methods

### Overview of Trial

The Pasos Hacia La Salud study was a randomized controlled trial of an internet-based PA intervention versus a wellness contact control in Latinas (N=205). Engagement data were collected between 2011 and 2014 and included information on website use (number of log-ins, time spent on websites, and number of times PA goals were set, personal PA reports were accessed, and PA tips features were used throughout the project period). Minutes of moderate to vigorous physical activity (MVPA) were measured subjectively and objectively at baseline, 6 months, and 12 months.

### Setting and Sample

The study was conducted at the University of California, San Diego, and human subjects approval was obtained from the institutional review board. The study was registered as a clinical trial [NCT01834287]. Participants in the trial included underactive Latinas (defined as participating in MVPA less than 60 minutes per week) aged between 18 and 65 years with regular internet access through home, work, or their community. Individuals were excluded from participation if they had any serious medical condition that would make unsupervised PA unsafe or were unable to read or speak Spanish fluently and demonstrate adequate functional health literacy (scoring at least a 17 on the Short Test of Functional Health Literacy in Adults [12]; currently pregnant or planning to become pregnant in the next year; planning to move from the area within the next year; hospitalized due to a psychiatric disorder in the past 3 years; or taking medication that may impair PA tolerance or performance [13]. The setting, sample, and primary outcomes are described in further detail in previously published manuscripts [13,14].

## Protocol

Participants were screened for initial eligibility by phone and then completed baseline PA assessments before being randomly assigned to one of two Spanish language internet-based conditions: Tailored Physical Activity Internet Intervention or Wellness Contact Control Internet Intervention. A detailed description of study protocols can be found elsewhere [13]. At the baseline randomization visit, research staff explained the intervention and participant expectations and helped intervention arm participants set realistic PA goals, identify potential barriers and potential solutions, and learn to use website features.

### *Tailored Physical Activity Internet Intervention*

The Tailored Physical Activity Internet Intervention was based on the transtheoretical model [15] and social cognitive theory [16]. Participants in this arm received access to the intervention website and completed monthly online surveys that generated automated tailored PA feedback reports on relevant theoretical constructs such as current stage of motivational readiness for PA, self-efficacy, and processes of change, as well as normative and progress feedback (ie, how they compared on these variables vs others who were already meeting guidelines and to their own prior responses) and useful facts on PA health benefits, stretching, and heart rate monitoring. The reports drew from a bank of more than 300 messages addressing different levels of these psychosocial and environmental factors affecting PA. In addition, participants received an online manual that was matched to their motivational readiness for PA. The manual emphasized strategies for increasing PA such as goal-setting, self-monitoring, problem-solving barriers, methods for increasing social support, and rewarding oneself for meeting PA goals.

Additional website features included (1) self-monitoring of minutes of PA and steps, (2) goal setting with graphs to compare goals to recorded minutes, (3) a message board to foster social support between participants, (4) an "ask the expert" section where participants could anonymously ask questions to a PhD-level researcher, and (5) online resources such as free exercise videos and maps to create walking routes. The intervention group received email prompts to access the intervention website weekly during month 1, biweekly during months 2 and 3, and monthly during months 4 to 6, with new PA information tip sheets made available on this schedule. Participants received monetary incentives to complete the study requirements, including $10 each month for completing the online monthly questionnaires. They also received a pedometer to track their steps (although minutes of MVPA rather than step count was the outcome of interest).

### *Wellness Contact Control Internet Group*

The Wellness Contact Control Internet Group received access to a Spanish language website with information on health topics other than PA. The Web-based content focused on diet and other factors associated with cardiovascular disease risk and included information from a series on heart health developed for Latinos by the National Heart Lung and Blood Institute. Control arm participants received the same number of email contacts on the same schedule as the intervention arm and also completed monthly surveys on wellness topics for the same $10 incentive offered to intervention arm participants.

## Measures

Demographics were assessed at baseline with a brief questionnaire assessing age, education, income, occupation, race, ethnicity, history of residence in the United States, marital status, and acculturation.

PA was measured subjectively using the 7-Day Physical Activity Recall (7-Day PAR) and objectively using accelerometers. The 7-Day PAR is an interviewer-administered instrument that provides an estimate of weekly minutes of PA and has consistently demonstrated acceptable reliability, internal consistency, and concurrent validity with objective measures of activity [17-19]. Accelerometers measure both movement and intensity of activity and have been validated with heart rate telemetry [20] and total energy expenditure [21]. Participants were asked to wear the accelerometer on their left hip for 7 days. Valid wear time was classified as 5 days of at least 600 minutes of wear time each day or at least 3000 minutes of wear time over 4 days. To be counted in the total minutes per week of MVPA, activity had to occur in ≥10-minute bouts, per the national PA guidelines at the time of the study. Accelerometer data was processed using the ActiLife software, with the established cut point of 1952 counts per minute to meet the minimum threshold for moderate intensity activity [22].

Website use was tracked throughout the project period using built-in software. Variables of interest included number of log-ins and time spent on websites for all participants, as well as the number of times key intervention website features related to PA goal-setting, personal PA reports, PA tips, and message board for social support were used. Self-reported satisfaction with the website was measured using consumer satisfaction questions on the follow-up surveys.

## Data Analysis

Overall log-ins to the study website were summarized monthly and compared between groups at each month using *t* tests. In addition, changes over time in number of log-ins within group was compared using a generalized linear model. Total time spent on the website was calculated for each arm during the intervention period (baseline to 6 months), maintenance period (6 to 12 months), and total study period (baseline to 12 months). Using a series of generalized linear models, we tested the effect of total time spent on the website and total number of log-ins over 12 months on both primary measures of intervention success (minutes per week of MVPA collected subjectively via the 7-day PAR and objectively via accelerometer). Models were adjusted for baseline MVPA, group, and wear time (in the case of accelerometry).

As a secondary outcome, we examined the effects of time spent on the website and number of log-ins on the odds of meeting national American College of Sports Medicine (ACSM) guidelines [23] for MVPA (≥150 minutes per week) at 12 months, measured subjectively and objectively.

Using a series of generalized linear models, we tested the association between time using each feature and PA outcomes (self-report and objectively measured minutes per week of MVPA) using a series of univariate models. Significant features (defined as those for which increases in use were significantly associated with increased minutes of MVPA at 12 months controlling for baseline) were then considered as part of a multivariate model predicting 12-month outcomes.

Associations between log-ins and targeted psychosocial constructs during the adoption phase were explored using a series of generalized linear models in which the mean value of the construct (eg, self-efficacy) at 6 months (primary end point) was regressed on baseline value of the construct, number of log-ins over 6 months, and treatment arm.

Finally, self-reported satisfaction with the website was summarized, and descriptive statistics are reported for the intervention arm. All analyses were carried out in SAS 9.3 (SAS Institute), with significance level set at alpha=.05 a priori.

## Results

Participants (N=205) were all Latina, mostly Mexican American (172/205, 83.9%), with an average age of 39.2 (SD 10.5) years. Full descriptions of participant demographics are presented in Table 1.

### Physical Activity Outcomes

Main outcomes of self-reported and objectively measured minutes of MVPA in both groups across all time points are reported in Table 2. Intervention arm participants demonstrated significantly greater gains in MPVA than control arm participants. Intervention participants were more likely to report meeting ACSM guidelines (≥150 minutes per week of MVPA) than control participants at 6 (31% vs 12%) and 12 (29% vs 19%) months, although group differences were not significant with accelerometer data at 6 (13% vs 9%) or 12 (16% vs 13%) months.

**Table 1.** Demographic characteristics.

| Characteristics | Intervention (n=104) | Control (n=101) |
|---|---|---|
| Hispanic, n (%) | 104 (100) | 101 (100) |
| Age (years), mean (SD) | 38.8 (10.6) | 39.6 (10.4) |
| First generation in the United States, n (%) | 90 (86.5) | 78 (77.2) |
| Body mass index (kg/m$^2$), mean (SD) | 29.1 (5.8) | 28.6 (4.5) |
| **Race, n (%)** | | |
| White | 47 (45.2) | 59 (58.4) |
| Mixed | 18 (17.3) | 15 (14.9) |
| Other | 32 (30.8) | 19 (18.8) |
| **Ethnicity, n (%)** | | |
| Mexican | 86 (82.7) | 87 (86.1) |
| Columbian | 2 (1.9) | 5 (5.0) |
| Guatemalan | 2 (1.9) | 0 (0) |
| Puerto Rican | 1 (1.0) | 1 (1.0) |
| Dominican Republic | 1 (1.0) | 0 (0) |
| Other | 15 (14.4) | 11 (10.9) |
| **Yearly household income, n (%)** | | |
| Less than $30,000 | 72 (69.3) | 64 (63.5) |
| $30,000-$50,000 | 18 (17.3) | 25 (24.7) |
| $50,000 or more | 10 (9.6) | 7 (6.9) |
| Don’t know | 4 (3.8) | 5 (5.0) |
| **Employment status, n (%)** | | |
| Unemployed | 51 (49.0) | 41 (41.0) |
| Part-time | 26 (25.0) | 30 (30.0) |
| Full-time | 26 (25.0) | 29 (29.0) |
| Refused/no answer | 1 (1.0) | 0 (0) |
| **Education level, n (%)** | | |
| Less than high school | 15 (14.6) | 14 (13.9) |
| High school graduate | 16 (15.5) | 8 (7.9) |
| Vocational/technical | 15 (14.6) | 12 (11.9) |
| Some college | 57 (55.4) | 67 (66.3) |
| **Language spoken in home, n (%)** | | |
| Only Spanish | 42 (40.4) | 35 (34.7) |
| More Spanish | 32 (30.8) | 33 (32.7) |
| Both equally | 16 (15.4) | 24 (23.8) |
| More English | 12 (11.5) | 5 (5.0) |
| Only English | 2 (1.9) | 4 (4.0) |
| **Marital status, n (%)** | | |
| Married | 52 (50.0) | 58 (57.4) |
| Living with partner | 5 (4.8) | 6 (5.9) |
| Separated | 14 (13.5) | 3 (3.0) |
| Divorced | 11 (10.6) | 17 (16.8) |
| Widowed | 2 (1.9) | 3 (3.0) |

| Characteristics | Intervention (n=104) | Control (n=101) |
|---|---|---|
| Never married | 20 (19.2) | 14 (13.9) |
| Health literacy score (score of 23-26 adequate) | 34.8 (2.7) | 37.2 (22.8) |

**Table 2.** Self-reported and objective moderate to vigorous physical activity at each study time point.

| Variable | Intervention, mean (SD) | | | Control, mean (SD) | | |
|---|---|---|---|---|---|---|
| | Baseline | 6 months | 12 months | Baseline | 6 months | 12 months |
| Self-reported MVPA[a] (min/wk), 7-day PAR[b], (n=205) | 8.0 (15.0) | 112.8 (97.1) | 108.6 (107.2) | 10.4 (24.0) | 63.0 (88.3) | 75.9 (89.8) |
| Accelerometer measured MVPA in 10-min bouts (min/wk), (n=200) | 35.8 (69.7) | 75.8 (91.0) | 70.4 (86.4) | 28.7 (48.2) | 43.0 (60.9) | 55.5 (74.6) |

[a]MVPA: moderate to vigorous physical activity.

[b]PAR: physical activity recall.

### Treatment Engagement

Overall, participants logged in to the website an average of 22 (SD 28) times over 12 months, with intervention participants logging on significantly more than controls (29 vs 14.7, *P*<.001). On average, intervention participants spent more time on the website at months 1, 4, and 6 compared to all other months, with maximum use at month 4. Unadjusted associations between log-ins and self-reported MVPA are presented in Figure 1. Unadjusted associations between log-ins and objectively measured MVPA are presented in Figure 2.

Both log-ins and time spent on the website were significantly related to intervention success (achieving higher mean minutes of MVPA per week at 12-month follow-up, controlling for baseline: *b*=.48, SE 0.20, *P*=.02 for objectively measured MVPA and *b*=.74, SE 0.34, *P*=.03 for self-reported MVPA). Furthermore, those meeting ACSM guidelines for PA at 12 months (≥150 minutes per week of self-reported MVPA) spent significantly more time on the website than those not meeting guidelines (35 vs 20 minutes over 12 months, *P*=.002). Figure 3 illustrates differences in the association between log-ins and self-reported MVPA by group over time. A significant difference in self-reported MVPA was apparent at 6 months for participants who logged in at least 1 time per month in the intervention arm (*P*=.04) but not the control arm (*P*=.54). At the 12-month follow-up, no difference was apparent in the intervention arm (*P*=.94), but the difference approached significance in the control arm (*P*=.054).

Among participants in the intervention arm, goal-setting features, personal PA reports, and PA tips were the most used portions of the website. Higher use of these features was associated with greater success in the program (more minutes of self-reported MVPA at 12 months controlling for baseline). Specifically, one additional use of these features per month over 12 months translated into an additional 34 minutes per week of MVPA (goals feature), 12 minutes per week (PA tips), and 42 minutes per week (PA reports). Results of the adjusted analyses are presented in Table 3.

Exploratory analyses suggest that independent of the effect of randomization, number of log-ins was positively associated with greater self-efficacy (*b*=.004, SE 0.002, *P*=.05) and social support (rewards and punishment subscale: *b*=.006, SE 0.003, *P*=.03) at 6 months such that more log-ins was associated with higher 6 months scores on these variables controlling for baseline. There were no significant effects of log-ins on psychosocial constructs (behavioral and cognitive processes, enjoyment, and social support family and friends scores) at 6 months.

**Figure 1.** Unadjusted association between log-ins and self-reported moderate to vigorous physical activity over 12 months.

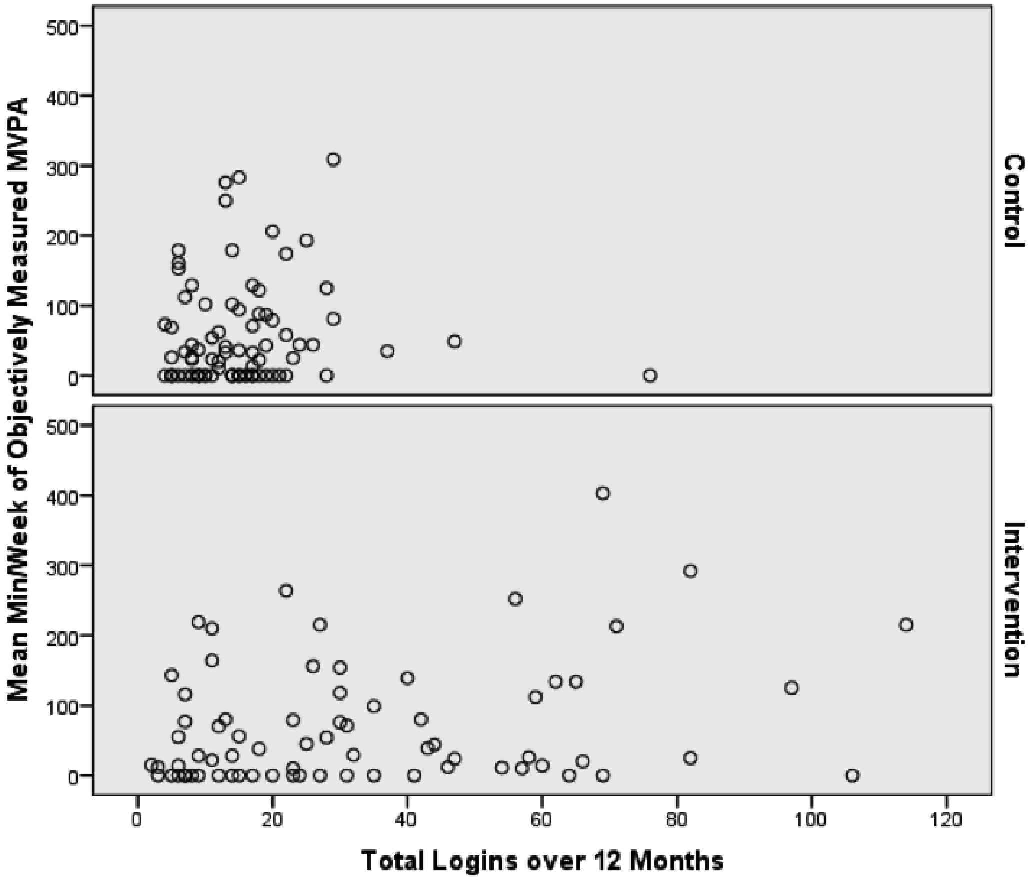



**Figure 2.** Unadjusted association between log-ins and objectively measured moderate to vigorous physical activity over 12 months.

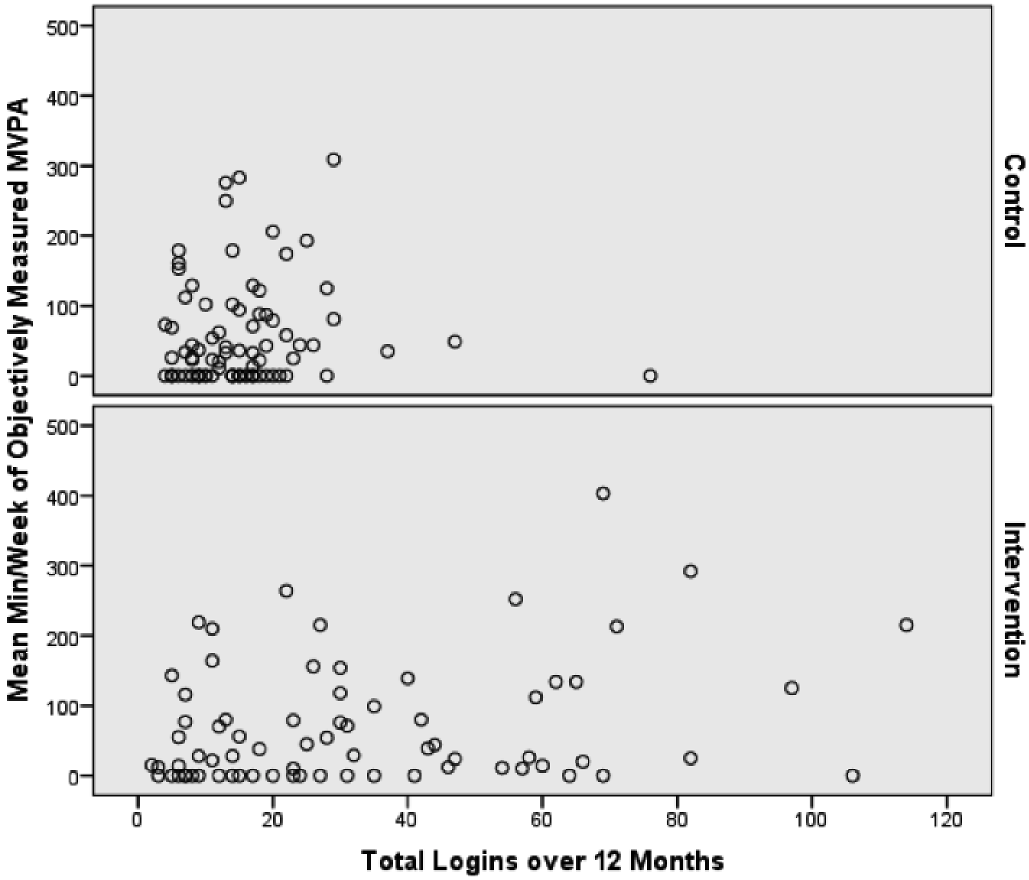


**Figure 3.** Comparing log-in rates and moderate to vigorous physical activity over time by group.

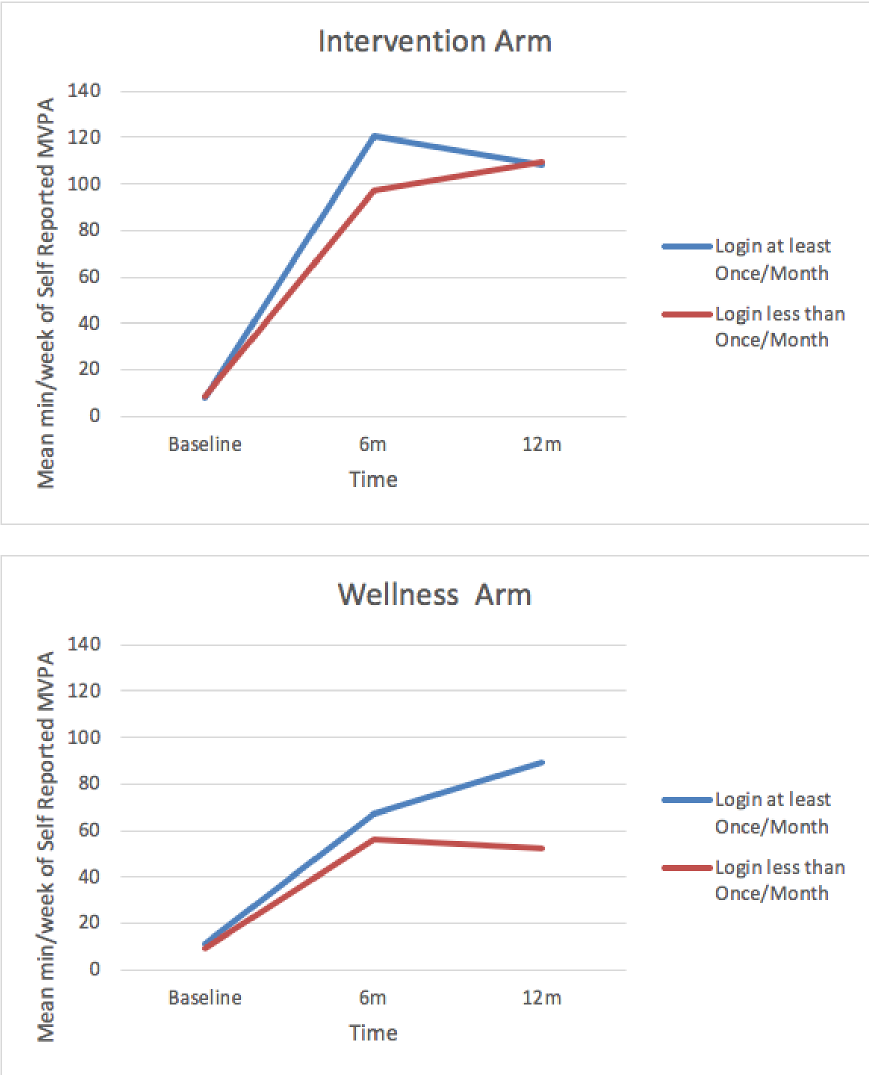



**Table 3.** Adjusted associations between website components and self-reported moderate to vigorous physical activity at 12 months among intervention participants.

| Component | *b* [a] | SE | *P* value |
|---|---|---|---|
| Goal setting | 2.85 | 1.38 | .04 |
| Physical activity tips | 1.00 | 0.82 | .05 |
| Physical activity reports | 3.49 | 1.28 | .008 |

[a]*b*: unstandardized regression coefficient. Models adjusted for baseline physical activity.

## Consumer Satisfaction Survey

All participants in the intervention arm reported that the intervention was at least somewhat enjoyable (66/66, 100%), and the majority (62/66, 94%) reported that the website was at least somewhat motivating. Ease of use of the website features was rated on a 5-point scale ranging from not at all easy (0) to very easy (5). The majority of participants reported that the following features were at least somewhat easy to use: recording activity (60/67, 90%), setting goals (59/67, 88%), message board (54/67, 81%), ask the expert (50/67, 75%), and places to be active (38/67, 57%). Less than half of participants (32/67, 48%) reported the MapMyWalk feature to be somewhat or very easy to use. Most participants reported using most of the features, including the self-monitoring/activity record (58/64, 91%), goal setting 58/64, 91%), message board (52/63, 83%), ask an expert (47/63, 75%), places to be active (37/63, 59%), and MapMyWalk (31/63, 49%). However, when asked to rank the most helpful feature of the website, the most commonly first ranked feature was MapMyWalk (10/63, 16%), followed by online exercise resources (6/63, 10%). The majority of participants reported gaining at least some knowledge from the website (64/66, 97%), and most (62/65, 95%) reported that the monthly email questionnaire reminders were at least somewhat helpful.

## Discussion

### Principal Findings

These results demonstrate that greater use of a tailored, internet-based PA intervention, particularly certain features on the website, was significantly related to increased PA levels in Latinas. These positive dose-response results support the existing treatment engagement literature [24-26] and suggest that PA and other health behavior interventions may be more successful than they appear in intent-to-treat analyses, which include all participants regardless of the actual dose they receive of the intervention. The results from this study are encouraging for health promotion researchers in that they demonstrate that an evidence-based internet-delivered intervention might have higher efficacy when used as intended. Given that lack of engagement is a major obstacle in health promotion research [27], these findings also reveal a need to identify effective strategies to promote engagement and treatment adherence.

Although participants of the Pasos Hacia La Salud intervention reported using most of the intervention features, our analyses of user data revealed that some features were more frequently used than others and that the level of engagement of participants with different intervention features had an effect on the number of minutes per week of MVPA. Engagement in internet-based health promotion interventions may be subject to unique challenges, driven mainly by limited face-to-face contact with participants [27,28]. Internet-based interventions, however, also grant unique opportunities for the implementation of strategies that have been associated with increased engagement, such as tailoring and personalization of messages [26,29,30]. In the Pasos Hacia La Salud intervention, tailoring and personalization were successfully employed in different website elements, including goal-setting and personalized reports; both of these intervention components were associated with the greatest increases in MVPA. On the other hand, increased use of other website features, including the possibility to interact with the researcher through the website and the group chat (meant to provide social support), was not associated with increased MVPA. In previous research, social support has produced mixed results for engagement, with some studies, but not all, showing increased support to be associated with increased engagement and adherence [30-32]. One study found that an online community increased intervention adherence and those who had little support at baseline used and benefited more from this feature compared with those who already had a supportive social network [31]. Perhaps a limitation of the message board feature on our website was that participants were anonymous and did not know each other, so they may not have been motivated to interact with each other or depend on each other for social support. Future research could consider including a social media component as part of the online intervention.

This study also revealed that overall engagement with the website peaked at key study time points (ie, around measurement and intervention visits) but overall decreased over time, which is also consistent with previous literature [24,25]. Moreover, the effect of increased engagement among intervention participants (specifically, logging in at least once per month) on minutes per week of MVPA seems to have dissipated during the maintenance period, as shown in Figure 3, perhaps as a result of this decreased engagement over time. A number of factors may help explain this decline in engagement. For example, participants may have become increasingly independent in their behavior and decreasingly dependent on tools available through the website or they may have become bored with the relatively static content despite regular updates such as daily PA tips and message board discussions. Another potential explanation is that these results may reflect the limited use of prompts and email reminders sent to participants during the tapered maintenance period (months 6 to 12) compared to the active intervention period (months 1 to 6), as previous research suggests that reminders are important to maintain engagement among intervention participants [32]. Curiously, the relationship between monthly log-ins and MVPA minutes in the control group grew stronger at 12 months, nearing statistical significance. Unfortunately we do not have an explanation for this unexpected differential finding. Identifying strategies for continuous engagement with behavior change tools during maintenance stages is an important area for future research. Text messages would be a possible intervention to explore.

Another important research question is to examine whether individuals and populations who are less tech-savvy are less likely to remain engaged with the different features of technology-based interventions. For example, it is notable that the feature MapMyWalk was ranked the most helpful by many intervention participants but was also seen as a less user-friendly feature by approximately half of the participants. Although this feature might have filled a specific need among those who were able to use it, increased user-friendliness might have enabled more participants to benefit from it. Thus, it is important to identify strategies to simplify this and similar intervention features to make them more accessible to participants who may be less tech-savvy in order to promote increased engagement and ensuing behavior change. Furthermore, future research should test interface design features that might better engage Latino populations so they are more likely to access and effectively use internet-based interventions for improving physical activity and other health behaviors. For example, future research could include delivery of internet- or app-based interventions that can be accessed on mobile phones rather than computers, as Latinos are heavily reliant on mobile phones for their internet access, more than other ethnicities. While Latinos have lagged other groups in accessing the internet and having broadband at home, they have been among the most likely to own a mobile phone and access the internet from a mobile device.

The results of this study are particularly encouraging for researchers focusing on Latino populations, who face a higher risk of chronic diseases and other health problems related to insufficient PA. Linguistically and culturally adapting existing evidence-based treatment materials to match the needs of this underserved population and delivering them via the internet appears to be a feasible way to reach this population. Overall, the majority of intervention participants reported some knowledge gain and found the intervention to be at least somewhat enjoyable and engaging. Additionally, successfully engaging participants in online tools to address behavior change techniques appears to result in significant increases in PA. Moreover, the frequency and time necessary for these significant gains was relatively minimal, suggesting that the participant burden was low and the cost was effective.

### Limitations

This study has certain limitations. Mainly, adherence and engagement cannot be experimentally tested, and thus we cannot rule out confounding variables driving the observed relationships. Other factors (eg, self-efficacy, motivation to change) may be driving both engagement and changes in MVPA. Additionally, although we have objective measures of engagement, such as number of log-ins and time spent on the website, these measures may be susceptible to error. For example, time spent on the website may be imprecise if participants remained logged into the website for a period of time but did not engage with it throughout the entire period (for example, if they were browsing other websites while remaining logged in to the study website). Nevertheless, by using two different objective measures of engagement (log-ins and time spent), as opposed to solely one measure, we aimed to corroborate the results and strengthen the evidence. Another limitation is that a component of the study intervention was emailing participants to remind them to log in and complete their monthly questionnaire in order to receive a $10 incentive, which may have differentially influenced the amount of user engagement with different features of the website, specifically encouraging use of the personalized reports feature. However, it may have also served to drive participants to the website when they would not have otherwise logged in at all and increased log-ins and time spent on multiple features of the website. As such incentives may not be replicable in dissemination of Web-based interventions, future research should examine the role and importance of incentives in website engagement. Monetary and nonmonetary (eg, attention, social support) incentives to participate in this study are not scalable in real-world settings, and thus the generalizability of the results from a public health standpoint should be tempered.

### Conclusions

Overall, results of this study suggested that greater use of an internet-based PA intervention, particularly certain features of the website, was significantly related to increased PA levels in Latinas. These results are encouraging from a public health perspective because this type of intervention can be delivered on a large scale at a relatively low cost once it is developed and published online. Because data suggests that Latinas are using the internet and other technology-based devices at a rapid pace, developing and disseminating these types of interventions is a potentially cost-effective way to help address the PA-related health disparities faced by this population. These findings also emphasize the importance of identifying effective ways to promote engagement and adherence to specific components to help ensure that participants receive the intended dose of the intervention.

## Acknowledgments

This work was supported by the National Cancer Institute of the National Institutes of Health (R01 CA159954). We would like to thank Raul Fortunet, Karla Nuñez, Rachelle Edgar, Madison Noble, Daniah Tanori, David Bakal, and Dr Veronica Villarreal at the University of California, San Diego, for their valuable research assistance and contributions to this study.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 3MB - jmir_v21i7e13063_app1.pdf ]

## References

1. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Schiller JS, Lucas JW, Ward BW, Peregoy JA. Summary health statistics for U.S. adults: National Health Interview Survey, 2010. Vital Health Stat 10 2012 Jan(252):1-207 [FREE Full text] [Medline: 22834228]
3. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
4. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
5. Jelalian E, Hart CN, Mehlenbeck RS, Lloyd-Richardson EE, Kaplan JD, Flynn-O'Brien KT, et al. Predictors of attrition and weight loss in an adolescent weight control program. Obesity (Silver Spring) 2008 Jun;16(6):1318-1323 [FREE Full text] [doi: 10.1038/oby.2008.51] [Medline: 18356834]

6. Hardcastle S, Blake N, Hagger MS. The effectiveness of a motivational interviewing primary-care based intervention on physical activity and predictors of change in a disadvantaged community. J Behav Med 2012 Jun;35(3):318-333. [doi: 10.1007/s10865-012-9417-1] [Medline: 22476812]
7. Leslie E, Marshall AL, Owen N, Bauman A. Engagement and retention of participants in a physical activity website. Prev Med 2005 Jan;40(1):54-59. [doi: 10.1016/j.ypmed.2004.05.002] [Medline: 15530581]
8. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
9. Lopez M, Gonzalez-Barrera A, Patten E. Closing the digital divide: Latinos and technology adoption. Washington: Pew Internet and American Life Project; 2013. URL: https://www.pewresearch.org/wp-content/uploads/sites/5/2013/03/Latinos_Social_Media_and_Mobile_Tech_03-2013_final.pdf [accessed 2019-06-18]
10. Brown A, López G, Lopez M. Digital divide narrows for Latinos as more Spanish speakers and immigrants go online. Washington: Pew Internet and American Life Project; 2016. URL: https://www.pewhispanic.org/wp-content/uploads/sites/5/2016/07/PH_2016.07.21_Broadbank_Final.pdf [accessed 2019-06-18]
11. Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: National Health Interview Survey, 2012. Vital Health Stat 10 2014 Feb(260):1-161 [FREE Full text] [Medline: 24819891]
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. [Medline: 14528569]
13. Marcus BH, Hartman SJ, Pekmezi D, Dunsiger SI, Linke S, Marquez B, et al. Using interactive Internet technology to promote physical activity in Latinas: Rationale, design, and baseline findings of Pasos Hacia La Salud. Contemp Clin Trials 2015 Aug 5:1. [doi: 10.1016/j.cct.2015.08.004] [Medline: 26255237]
14. Marcus BH, Hartman SJ, Larsen BA, Pekmezi D, Dunsiger SI, Linke S, et al. Pasos Hacia La Salud: a randomized controlled trial of an internet-delivered physical activity intervention for Latinas. Int J Behav Nutr Phys Act 2016;13(1):62 [FREE Full text] [doi: 10.1186/s12966-016-0385-7] [Medline: 27234302]
15. Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. Res Q Exerc Sport 1992 Mar;63(1):60-66. [doi: 10.1080/02701367.1992.10607557] [Medline: 1574662]
16. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs: Prentice-Hall; 1986.
17. Richardson MT, Ainsworth BE, Jacobs DR, Leon AS. Validation of the Stanford 7-day recall to assess habitual physical activity. Ann Epidemiol 2001 Feb;11(2):145-153. [Medline: 11164131]
18. Sallis JF, Haskell WL, Wood PD, Fortmann SP, Rogers T, Blair SN, et al. Physical activity assessment methodology in the Five-City Project. Am J Epidemiol 1985 Jan;121(1):91-106. [Medline: 3964995]
19. Hayden-Wade HA, Coleman KJ, Sallis JF, Armstrong C. Validation of the telephone and in-person interview versions of the 7-day PAR. Med Sci Sports Exerc 2003 May;35(5):801-809. [doi: 10.1249/01.MSS.0000064941.43869.4E] [Medline: 12750590]
20. Janz KF. Validation of the CSA accelerometer for assessing children's physical activity. Med Sci Sports Exerc 1994 Mar;26(3):369-375. [Medline: 8183103]
21. Melanson EL, Freedson PS. Validity of the Computer Science and Applications Inc (CSA) activity monitor. Med Sci Sports Exerc 1995 Jun;27(6):934-940. [Medline: 7658958]
22. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications Inc accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [Medline: 9588623]
23. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
24. 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):84-107 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
25. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011;13(1):70-83 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
26. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
27. 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. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
28. Leslie E, Marshall AL, Owen N, Bauman A. Engagement and retention of participants in a physical activity website. Prev Med 2005 Jan;40(1):54-59. [doi: 10.1016/j.ypmed.2004.05.002] [Medline: 15530581]

29. Morrison L, Moss-Morris R, Michie S, Yardley L. Optimizing engagement with Internet-based health behaviour change interventions: comparison of self-assessment with and without tailored feedback using a mixed methods approach. Br J Health Psychol 2014 Nov;19(4):839-855 [FREE Full text] [doi: 10.1111/bjhp.12083] [Medline: 24308806]
30. Redfern J, Santo K, Coorey G, Thakkar J, Hackett M, Thiagalingam A, et al. Factors influencing engagement, perceived usefulness and behavioral mechanisms associated with a text message support program. PLoS One 2016;11(10):e0163929 [FREE Full text] [doi: 10.1371/journal.pone.0163929] [Medline: 27741244]
31. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010;12(4):138-153 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
32. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen J. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):17-40 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]


## Abbreviations

**PA:** physical activity
**MVPA:** moderate to vigorous physical activity
**PAR:** Physical Activity Recall
**ACSM:** American College of Sports Medicine


*Edited by G Eysenbach; submitted 09.12.18; peer-reviewed by R Buys, E Guisado-Fernandez; comments to author 07.02.19; revised version received 08.06.19; accepted 10.06.19; published 24.07.19.*

*Please cite as:*
*Linke SE, Dunsiger SI, Gans KM, Hartman SJ, Pekmezi D, Larsen BA, Mendoza-Vasconez AS, Marcus BH*
*Association Between Physical Activity Intervention Website Use and Physical Activity Levels Among Spanish-Speaking Latinas: Randomized Controlled Trial*
*J Med Internet Res 2019;21(7):e13063*
*URL: http://www.jmir.org/2019/7/e13063/*
*doi:10.2196/13063*
*PMID:31342902*

Original Paper

# Assessment of Medication Adherence Using a Medical App Among Patients With Multiple Sclerosis Treated With Interferon Beta-1b: Pilot Digital Observational Study (PROmyBETAapp)

Volker Limmroth[1], MD; Klaus Hechenbichler[2], PhD; Christian Müller[3], PhD; Markus Schürks[3], MSc, MD

[1]Clinic for Neurology and Palliative Medicine, Municipal Hospital Köln-Merheim, Cologne, Germany

[2]Institut Dr. Schauerte, Munich, Germany

[3]Bayer Vital GmbH, Leverkusen, Germany

**Corresponding Author:**
Markus Schürks, MSc, MD
Bayer Vital GmbH
Neurologie, Immunologie & Ophthalmologie
Leverkusen, 51368
Germany
Phone: 49 2143051455
Email: markus.schuerks@bayer.com

## *Abstract*

**Background:** Accurate measurement of medication adherence using classical observational studies typically depends on patient self-reporting and is often costly and slow. In contrast, digital observational studies that collect data directly from the patient may pose minimal burden to patients while facilitating accurate, timely, and cost-efficient collection of real-world data. In Germany, ~80% of patients with multiple sclerosis (MS) treated with interferon beta 1b (Betaferon) use an electronic autoinjector (BETACONNECT), which automatically records every injection. Patients may also choose to use a medical app (myBETAapp) to document injection data and their well-being (using a “wellness tracker” feature).

**Objective:** The goal of this pilot study was to establish a digital study process that allows the collection of medication usage data and to assess medication usage among patients with MS treated with interferon beta-1b who use myBETAapp.

**Methods:** The PROmyBETAapp digital observational study was a mixed prospective and retrospective, noninterventional, cohort study conducted among users of myBETAapp in Germany (as of December 2017: registered accounts N=1334; actively used accounts N=522). Between September and December 2017, users received two invitations on their app asking them to participate. Interested patients were provided detailed information and completed an electronic consent process. Data from consenting patients’ devices were collected retrospectively starting from the first day of usage if historical data were available in the database and collected prospectively following consent attainment. In total, 6 months of data on medication usage behavior were collected along with 3 months of wellness tracker data. Descriptive statistics were used to analyze persistence, compliance, and adherence to therapy.

**Results:** Of the 1334 registered accounts, 96 patients (7.2%) provided informed consent to participate in the study. Of these, one patient withdrew consent later. For another patient, injection data could not be recorded during the study period. Follow-up of the remaining 94 patients ended in May 2018. The mean age of participants was 46.6 years, and 50 (53%) were female. Over the 6-month study period, persistence with myBETAapp usage was 96% (90/94), mean compliance was 94% of injections completed, and adherence (persistence and ≥80% compliance) was 89% (84/94). There was no apparent difference between male and female participants and no trend across age groups. The wellness tracker was used by 21% of participants (20/94), with a mean of 3.1 entries per user.

**Conclusions:** This study provides important information on medication usage among patients with MS treated with interferon beta-1b and on consenting behavior of patients in digital studies. In future studies, this approach may allow patients’ feedback to be rapidly implemented in existing digital solutions.

**Trial Registration:** ClinicalTrials.gov NCT03134573; https://clinicaltrials.gov/ct2/show/NCT03134573

***(J Med Internet Res 2019;21(7):e14373)*** doi:10.2196/14373

**KEYWORDS**

digital observational study; BETACONNECT; myBETAapp; interferon beta-1b; multiple sclerosis; medication adherence; medication compliance; medication persistence

## Introduction

### Background

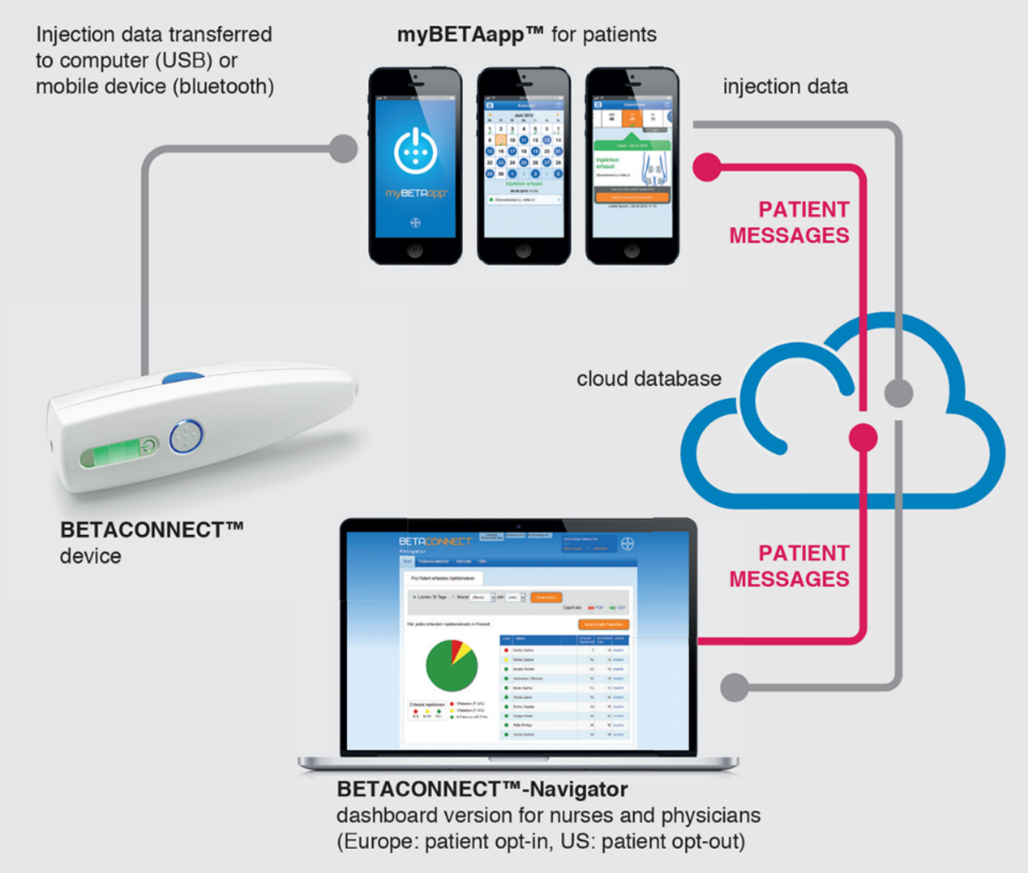
Inadequate medication usage is a major challenge in all diseases requiring long-term treatment [1]. Although adherence is typically high during the treatment of acute conditions, for chronic diseases, it decreases dramatically after the first 6 months of therapy [2]. Despite the importance of adherence to therapy, assessing medication usage in real-world settings is challenging. Large-scale retrospective studies are typically conducted using prescription data, but may not accurately measure adherence to treatment regimens [3]. Furthermore, accurate methods of measuring medication adherence using classical observational studies typically depend on patient self-reporting and are often costly and slow. In particular, longitudinal studies, needed to assess adherence over time, are more costly than cross-sectional studies or database analyses [4,5], and repeated clinic visits can be a burden for both clinicians and patients [6]. In contrast, digital observational studies that collect data directly from the patient may pose minimal burden to patients while facilitating accurate, timely, and cost-efficient collection of real-world data on medication usage.

One chronic disease for which adherence to medication is important is multiple sclerosis (MS), a chronic autoimmune disease of the central nervous system, which typically starts in young adulthood. The most common subtype of MS is relapsing–remitting MS, which is characterized by episodes of neurological dysfunction (relapses) separated by periods of remission and recovery [7]. There is no cure for MS, and effective disease management requires strict adherence to treatment regimen dose and administration schedules, typically involving disease-modifying drugs (DMDs) [8-11]. However, 25%-50% of patients with MS taking DMDs are not adherent [12,13]; among patients using injectable DMDs, the most common reason for nonadherence is forgetting to administer the injection [12]. Medication adherence among patients with MS has a direct effect on treatment outcomes, including relapse rates and health-related quality of life [8,9], as well as on health care resource utilization and costs [8-10].

In Germany, ~80% of patients with MS treated with interferon beta-1b (Betaferon) use an electronic autoinjector (BETACONNECT; Bayer, Leverkusen, Germany), which automatically records every injection. Patients may also choose to use a medical app (myBETAapp) [14] to document injection data, which can be automatically transferred from the electronic autoinjector or entered manually. Patients can also use the app to document their wellbeing using the “wellness tracker” feature, which allows them to manually record the following data on a Likert scale: ability to walk, coordination, energy level, bladder control, exercise level, memory, vision, bowel control, emotions, and eating habits. In addition, for patients who have first provided informed consent for their data to be stored in the myBETAapp database, whenever their smartphone is connected to the internet, the injection-related data and the wellness-related data are transferred to an external server under the surveillance of an external host. If patients agree (by signing a second electronic informed consent form together with their treating health care provider), data can be shared with their health care provider via an independently hosted online database (Figure 1). The three components of the BETACONNECT system—the autoinjector, myBETAapp, and the BETACONNECT Navigator—constitute an ecosystem aimed at supporting patients with MS.

Previous studies have shown that electronic autoinjector use is associated with a high level of adherence and patient satisfaction [15,16]. In addition, most patients (70/75) using myBETAapp find it helpful for regular injections, and approximately half (34/75) use the data sharing feature [17].

**Figure 1.** Components of the BETACONNECT system. Republished with permission from Future Medicine Ltd, from Limmroth et al, 2018 [17]; permission conveyed through Copyright Clearance Center, Inc. BETACONNECT Navigator was not part of the PROmyBETAapp study.




### Objectives

The widespread use of the electronic autoinjector in Germany makes it an ideal setting for a digital observational study. The goals of this pilot study were to establish a digital study process that allows the collection of medication usage data and to investigate medication usage among patients treated with interferon beta-1b using myBETAapp. In addition, we aimed to investigate the proportion of patients consenting to participate in the study, and the proportion willing to use the wellness tracker.

## Methods

### Study Design and Patients

The PROmyBETAapp digital observational study (trial registration: NCT03134573 [18]) was a mixed prospective and retrospective, noninterventional, cohort study conducted among users of myBETAapp in Germany (as of December 2017, registered accounts: N=1334, actively used accounts: N=522; active usage of an account is defined as having actively added, deleted, or changed data in myBETAapp during the previous month). Adult patients with MS treated with interferon beta-1b were eligible to participate in the study if they were using the app and provided electronic informed consent.

### Study Conduct and Ethics Approval

Between September and December 2017, users received two invitations on their app asking them to participate. Patients who expressed interest in the study (by pressing a button) were presented with more detailed information in a sequential manner: background, aim of the study, study design and data usage, data privacy including how to give and withdraw consent, and contact information at the database host and Bayer in case of questions. After each sequence, patients were required to confirm that they had understood the information and were finally asked to give their consent to participate. The patient informed consent document included text requesting patients to report any side effects or possible side effects to their physicians or nurses, or directly to the Bayer pharmacovigilance department. Only patients consenting to all steps were able to participate in the study.

The study protocol was approved by the ethics committee of the Nordrhein Medical Chamber (approval number AZ 2017170).

In this observational study, interferon beta-1b was prescribed in accordance with the terms of the marketing authorization. There was no assignment of a patient to a particular therapeutic strategy. The treatment decision fell within current practice, and the prescription of the medicines was clearly separated from the decision to include the patient in the study. No additional diagnostic or monitoring process was required for enrolment or during the study. Furthermore, patients made their decisions to use myBETAapp and participate in the study freely. In addition, patients were free to withdraw from the study at any time and without giving a reason; after withdrawal of a patient from the study, data from that patient were not used for any further analyses. No investigator was involved in the patient recruitment or data collection processes.

## Data Collection

Data from consenting patients' devices were collected retrospectively, starting from the first day of usage if historical data were available in the database and collected prospectively following consent attainment. In total, 6 months of data on medication usage behavior were collected along with 3 months of wellness tracker data. With respect to the wellness tracker, the only data recorded were whether a participant used it and, if so, how often. All data collected in myBETAapp were transferred via the internet to a database hosted by TWT Digital Health GmbH (Heidelberg, Germany). Data collected directly or calculated by TWT Digital Health GmbH based on directly collected data are listed in Table 1.

**Table 1.** Data derived from the electronic autoinjector and myBETAapp (directly collected or calculated) that were transferred to the study database.

| Data | Description |
|---|---|
| **BETACONNECT injection** | |
| patient_id | Patient identifier |
| injection_date_times-tamp_utc | Data recorded via BETACONNECT: date and time of injection within the first 6 months of usage |
| dose | Data recorded via BETACONNECT: dose of interferon beta-1b; possible values: 0.25, 0.5, 0.75, 1.00 |
| injection_site | Data recorded via BETACONNECT: "unknown"; manual data entry via myBETAapp |
| flag_manual_autoinjector | Autoinjector: data recorded via BETACONNECT; Manual: data manually entered by patient into myBETAapp |
| needle_depth | Data recorded via BETACONNECT: injection depth; possible values: 8, 10, 12 |
| injection_speed | Data recorded via BETACONNECT: speed of injection; possible values: low, medium, high |
| **Patient** | |
| patient_id | Patient identifier |
| acceptance_date_utc | Date of patient's consent to participate |
| first_injection_date_utc | Not used for analysis |
| age | Age of patient |
| gender | Gender of patient; possible values: M, F |
| complete_wellness_number | Number of completed wellness tracker entries within the first 3 months of usage |

## Data Quality

Data stored in the database were cleaned before analysis. Patients were free to choose not to download their injection data from the electronic autoinjector into the app, instead documenting injections manually or choosing not to record any injection information in the app. In addition, a number of factors could have affected the quality of the collected data, including data collection and synchronization of data from the electronic autoinjector at multiple time points, patient preferences for the myBETAapp reminder function, and settings for forwarding data from the app to the database. As a result, it was possible for the database to include multiple injection records for a given day. Therefore, prior to data analysis injection, data were automatically corrected according to the rules described in Textbox 1. Redundant data not considered by these rules were not cleaned manually. All data transferred to Bayer were anonymized.

**Textbox 1.** Data cleaning procedure.

**Identical data entries**

- Duplicate data entries were deleted.

**Multiple data entries on same day**

- If more than one autoinjector record or more than one manual data entry was available for the same day, the last entry (ordered by time) was kept.

**Combination of manual and autoinjector data on the same day**

- If both autoinjector and manual data entries were present for a given day, the autoinjector record was used.
- If the autoinjector record injection site was "unknown," the injection site from the manual record was substituted.
- If the autoinjector record included injection site data, the autoinjector record was used, even if the manual record included different injection site data.

## Statistical Analysis

### *Statistical Methods*

Analyses of persistence, compliance, and adherence, defined as shown in Textbox 2, were conducted using descriptive statistics. The analyses included a stratified analysis according to gender and age subgroups.

Missing data can result from various reasons, including technical issues and patients choosing not to document additional information on injections via the app.

**Textbox 2.** Study definitions.

- Persistence: The patient was still using interferon beta-1b at the end of the 6-month study period.
- Compliance: Percentage of expected doses actually injected.
- Adherence: The patient was both persistent and ≥80% compliant.

### *Study Size*

Injection-related data from 40 patients would allow determination of mean compliance (%) with ±10% for a two-sided approximate 95% CI with >99% confidence (assuming an SD of 15%-20%).

### Sensitivity Analysis

Sensitivity analyses were conducted on data derived from the BETACONNECT autoinjector only. A patient was included in the sensitivity analysis if he/she had provided only injection data recorded by the BETACONNECT throughout the observation period.

# Results

## Participants

Of the 1334 registered myBETAapp accounts, 96 patients (7.2%) provided their informed consent to participate in the study. Of these, one patient withdrew informed consent later. For another patient injection data could not be recorded during the study period, apparently because of a technical issue. Follow-up of the remaining 94 patients ended in May 2018. The mean age of participants was 46.6 years, 50 (53%) were female, and 44 (47%) were male (Table 2).

**Table 2.** Study participants according to their age (N=94).

| Age group (years) | Total, n (%) | Female, n (%) | Male, n (%) |
|---|---|---|---|
| <30 | 10 (11) | 9 (18) | 1 (2) |
| 30-39 | 23 (24) | 14 (28) | 9 (20) |
| 40-49 | 32 (34) | 15 (30) | 17 (39) |
| 50-59 | 15 (16) | 5 (10) | 10 (23) |
| ≥60 | 14 (15) | 7 (14) | 7 (16) |

## Data Collected

### *Injection Data*

For most patients (54/94, 57%), only autoinjector data were available, while 26 of 94 patients (28%) used both autoinjector data and manual documentation (Table 3). In total, 31 of the 50 female participants (62%) and 23 of the 44 male participants (52%) did not enter any manual data. In the group aged 30-39 years, only 8 of the 23 patients (35%) used autoinjector data alone, and 11 (48%) used both autoinjector data and manual documentation. The majority of participants (60/94, 64%) recorded injection location data.

**Table 3.** Injection and wellness tracker data collected (N=94).

| Data | Total | Female | Male |
|---|---|---|---|
| **Injection, n (%)** | | | |
| **Injection patterns** | | | |
| Autoinjector data only | 54 (57) | 31 (62) | 23 (52) |
| Manual data only | 14 (15) | 8 (16) | 6 (14) |
| Both autoinjector and manual data | 26 (28) | 11 (22) | 15 (34) |
| **Injection location data** | | | |
| Yes | 60 (64) | 31 (62) | 29 (66) |
| No | 34 (36) | 19 (38) | 15 (34) |
| **Wellness tracker** | | | |
| **Use of wellness tracker, n (%)** | | | |
| Yes | 20 (21) | 13 (26) | 7 (16) |
| No | 74 (79) | 37 (74) | 37 (84) |
| **Number of wellness tracker entries per user** | | | |
| Number of users | 20 | 13 | 7 |
| Mean (SD) | 3.1 (4.1) | 3.6 (5.1) | 2 (0.8) |
| Minimum | 1 | 1 | 1 |
| Q1[a] | 1 | 1 | 1 |
| Median | 2 | 1 | 2 |
| Q3[b] | 2.5 | 2 | 3 |
| Maximum | 17 | 17 | 3 |

[a]Q1: lower 25% quartile.

[b]Q3: upper 75% quartile.

### *Wellness Tracker Data*

Over the 3-month observation period, the wellness tracker was used by 20 of 94 participants (21%), with a mean of 3.1 entries per user (Table 3). Female patients were more likely to use the wellness tracker than male participants (13/50, 26%, vs 7/44, 16%) and made more entries (mean: 3.6 vs 2.0 entries per user).

## Medication Usage

Over the 6-month study period, persistence was 96%, with only 4 of 94 patients discontinuing treatment. The majority of patients did not miss any injections (median compliance: 100%, mean: 94%, Table 4 and Figure 2). Compliance was lowest among patients aged 30-39 years (median 100%, mean 89%). Adherence was 89% over the 6-month observation period (84/94 participants were adherent; Figure 2). There was no apparent difference between female (44/40, 88%) and male (40/44, 91%) participants and no trend across age groups.

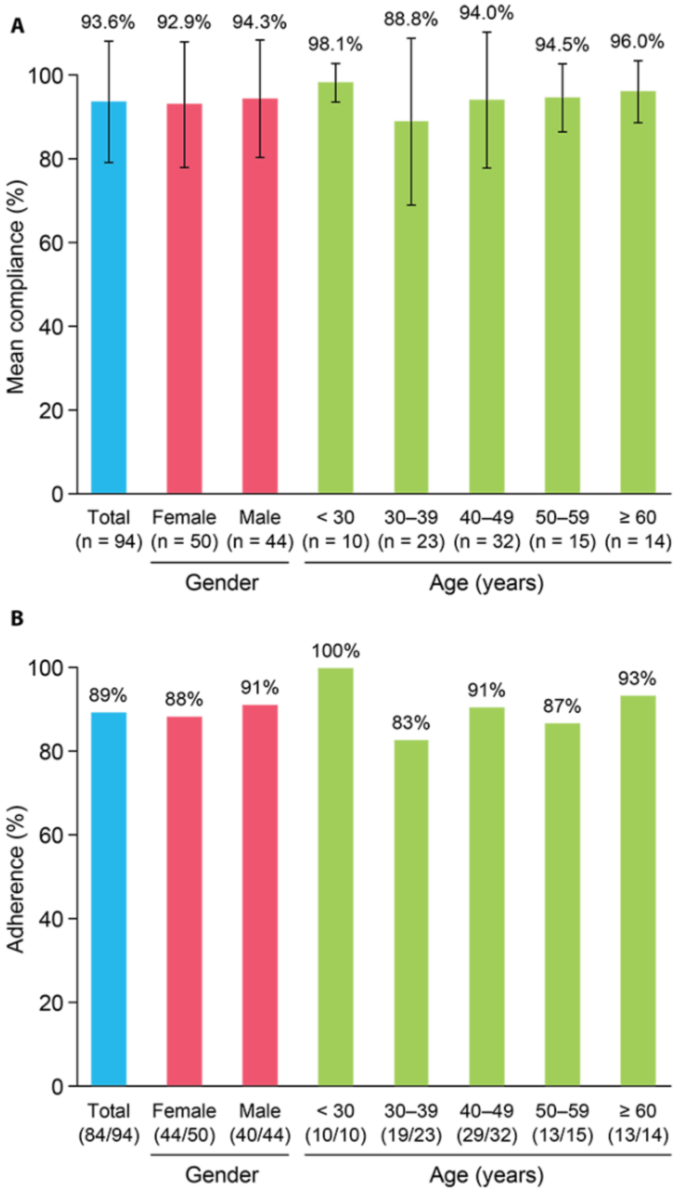
**Table 4.** Persistence, compliance, and adherence in the main analysis and sensitivity analysis groups.

| Data | Total | Female | Male |
|---|---|---|---|
| **Main analysis group** | | | |
| Participants, n (%) | 94 (100) | 50 (53) | 44 (47) |
| **Persistence, n (%)** | | | |
| Yes | 90 (96) | 47 (94) | 43 (98) |
| No | 4 (4) | 3 (6) | 1 (2) |
| **Compliance** | | | |
| Mean (SD) | 93.6 (14.5) | 92.9 (15) | 94.3 (14) |
| Minimum | 23.3 | 35.6 | 23.3 |
| Q1[a] | 94.4 | 94.4 | 95.6 |
| Median | 100 | 99.4 | 100 |
| Q3[b] | 100 | 100 | 100 |
| Maximum | 106.7 | 106.7 | 106.7 |
| **Adherence, n (%)** | | | |
| Yes | 84 (89) | 44 (88) | 40 (91) |
| No | 10 (11) | 6 (12) | 4 (9) |
| **Sensitivity analysis group** | | | |
| Participants, n (%) | 54 (100) | 31 (57) | 23 (43) |
| **Persistence, n (%)** | | | |
| Yes | 52 (96) | 29 (94) | 23 (100) |
| No | 2 (4) | 2 (6) | 0 (0) |
| **Compliance** | | | |
| Mean (SD) | 93.6 (13.6) | 92.8 (15.7) | 94.7 (10.5) |
| Minimum | 35.6 | 35.6 | 57.8 |
| Q1[a] | 95.6 | 95.6 | 94.4 |
| Median | 98.9 | 98.9 | 100 |
| Q3[b] | 100 | 100 | 100 |
| Maximum | 101.1 | 101.1 | 101.1 |
| **Adherence, n (%)** | | | |
| Yes | 49 (91) | 28 (90) | 21 (91) |
| No | 5 (9) | 3 (10) | 2 (9) |

[a]Q1: lower 25% quartile.

[b]Q3: upper 75% quartile.

**Figure 2.** (A) Mean compliance and (B) adherence over the 6-month study period. Bars in (A) indicate SD.




### Sensitivity Analysis

All analyses performed for the full analysis set (N=94) were repeated in the subgroup of patients for whom only autoinjector data were available (N=54; Table 4). The demographics of the sensitivity analysis group (mean age, 47.2 years; 31/54 or 57% female) were similar to those of the full analysis set (Table 2). Persistence (96% in both analyses [90/94 and 52/54 participants included in the full analysis and sensitivity analysis, respectively]), compliance (median: 100%, mean: 94% in full analysis; median: 99%, mean: 94% in sensitivity analysis), and adherence (84/94, 89% in full analysis vs 49/54, 91% in sensitivity analysis) were similar in the two groups.

## Discussion

### Principal Results

This pilot study demonstrates the feasibility of a digital observational study design, with recruitment, consent, and data collection conducted via myBETAapp, and data retrieved and processed in a time- and cost-efficient manner. The proportion of patients participating was 7.2% of the registered accounts. Injection data were successfully obtained for almost all participants, and the sensitivity analysis showed that useful data can be generated without any manual input from patients. Only a minority of patients chose to use the optional wellness tracker function. The results show that for patients with MS using the app in Germany, persistence, compliance, and adherence over the 6-month period were high.

### Comparison With Prior Work

Overall, the proportion of participants who were male (44/94, 47%) was higher than that seen in a recent German observational study of patients with MS using interferons or glatiramer acetate (120/429, 28%) [19]. This proportional shift may be due to the digital study design.

Adherence based on data collected through myBETAapp was slightly higher (84/94, 89%) than that obtained in a previous nondigital study of patients using the electronic autoinjector (62/77, 80.5% [at 24 weeks]) [15]. A potential explanation for this may be the reminder function implemented in myBETAapp; however, we did not investigate use of this feature in our study. No comparative data on persistence, compliance, or adherence between patients receiving interferon beta-1b who use the myBETAapp and those who do not use the app are available. A similar proportion of adherent patients (131/158, 82.9% [at 24 weeks]) was seen in another nondigital observational study of patients with MS using a different electronic injection system [20]. Therefore, medication usage data obtained through the app are likely to be reliable.

Consistent with previous studies on MS [8,21,22], in this analysis, compliance and adherence to medication were lowest among participants aged 30-39 years. For this group of working-age patients, the effectiveness of MS treatment may have a significant impact on long-term health-related quality of life. One strength of digital observational studies of this kind is the potential to rapidly distinguish patient cohorts for whom particular treatment regimens or support may be appropriate.

### Potential of Digital Observational Studies

Compared with classical observational approaches, digital studies have considerable potential to reduce costs and improve efficiency. The costs of conducting longitudinal observational studies can approach those of randomized controlled trials [5], primarily due to the costs associated with clinical sites. In particular, recruitment of a large patient population may necessitate the involvement of many clinical sites, leading to complexity and high costs; in the United States, site management costs make up more than half the cost of phase 4 trials [23]. To avoid site costs, a UK trial of 15,480 patients with diabetes (ASCEND) was conducted by mail, with overall costs an order of magnitude lower than those of traditional clinical studies [24]. However, recruitment in the ASCEND trial was slow [24], a common problem in clinical studies [25]. By enabling eligible patients to be approached rapidly outside of routine clinic visits, digital methods may enable studies to be completed more quickly than traditional studies and at a lower cost. In addition, once the digital platform has been established, the marginal cost of adding additional patients is small, meaning that digital observational studies can potentially enroll very large populations of patients. For example, the Apple Heart Study recently enrolled more than 400,000 participants in a 9-month period [26].

Many patients are happy to take part in clinical studies, both to potentially improve their own treatment and help others by contributing to scientific research [27]. However, participation may be limited for several reasons, including the inconvenience of having to attend additional clinic visits [28]. In addition, patients whose doctors are not investigators in clinical studies may not be offered the opportunity to participate [29]. Both of these barriers to clinical study participation can potentially be overcome by digital study designs, with patients being able to take part in studies remotely and potentially without the direct involvement of their treating physician. Appropriately designed digital recruitment processes may also increase patient engagement; directly approaching patients to participate may make them feel like their contribution is valued and important in a way that being selected by their physician would not. The US Food and Drug Administration’s newly developed MyStudies app is an example of a “platform” app that may facilitate the conduct of digital studies on a large scale [30].

### Limitations

This pilot study has several limitations. First, the data used in this study were obtained both prospectively (after informed consent) and retrospectively (before informed consent). Although retrospectively collected data in observational studies are prone to recall bias, we believe that this is not an issue in our study. Although some data were collected before the start of the study, they were collected in the same way as the prospective data, by automatic recording of injection data. However, we cannot exclude the possibility that participation in the PROmyBETAapp study may have changed the injection behavior of the participants. Second, the technology used to obtain data on medication intake behavior using the electronic autoinjector and myBETAapp has not been validated. However, there is no gold standard for recording medication intake in observational studies, and direct surveillance is not feasible in such a setting. Third, technical issues with the smartphone (eg, connection between the electronic autoinjector and the smartphone, or between the smartphone and the server) as well as patients’ decision not to document additional data may have led to missing data, which have not been replaced. Fourth, for the 4% of patients who were classified as nonpersistent, it is not possible to distinguish between patients discontinuing treatment and those simply ceasing to document their injections in the app. Fifth, only patients using interferon beta-1b participated in the study, limiting generalizability to patients using other DMDs. However, the aim of the study was to investigate medication intake behavior among patients using the app, and the electronic autoinjector is only available to patients using interferon beta-1b and not to those using other medications. Sixth, only 7.2% of the registered account users consented to participate in the study, which may further limit generalizability to all patients using interferon beta-1b. Seventh, it is possible that mainly technophile patients who were using myBETAapp decided to participate in the study, which may constitute a selection bias. However, MS predominantly affects young people, who tend to be familiar with using mobile devices and apps, and any bias introduced is likely to be limited. In addition, in this pilot study, the results for persistence, compliance, and adherence were similar among men and women and across age groups. Eighth, the results must be interpreted with caution due to the limited sample size, especially in the gender and age subgroups. Finally, results beyond 6 months are not yet available. However, studies using the same design over longer periods of time are planned.

### Conclusions

This study provides important information on medication usage and consenting behavior of patients in digital studies. Persistence, compliance, and adherence over a 6-month period were high for patients with MS using the app. There are some open questions, mainly regarding recruitment and study conduct, which need to be addressed in the future studies. For example, we need to develop more refined approaches to ensure that participating patients are representative of the whole population of interest. In this context, data privacy aspects may be important. Specifically, data privacy may determine patients' willingness to share data, in general (ie, to participate or not participate in a digital study). Willingness to share certain data may further differ by data type, with clinical and wellness data being potentially more sensitive than medication usage data. In addition, we need to use methodology that allows us to differentiate between patients terminating their medication and those simply terminating use of myBETAapp. In future studies, this approach may allow patients' feedback to be rapidly implemented into existing digital solutions. More comprehensive studies using the digital observational design will be conducted, investigating more clinical and patient-reported outcomes.

## Acknowledgments

We thank Dr Paul Overton of Beacon Medical Communications Ltd (Brighton, UK) for providing medical writing support, which was funded by Bayer Vital (Leverkusen, Germany) in accordance with Good Publication Practice (GPP3) guidelines.

## Authors' Contributions

CM and MS designed the study, with medical and scientific input from VL. CM conducted the study, and KH carried out the statistical analysis. KH and MS interpreted the data. All authors contributed to the drafting of the article and revised the manuscript drafts for important intellectual content. All authors have reviewed and approved the final manuscript.

## Conflicts of Interest

VL has served as advisor or speaker or received research grants from Antisense, Allergan, Bayer, Biogen, Genzyme, Novartis, Roche, and TEVA. CM and MS are full-time employees of Bayer Vital GmbH (Leverkusen, Germany). KH is an employee of the Institute Dr. Schauerte (Munich, Germany). Bayer selected the Institute Dr. Schauerte for statistical analysis of the PROmyBETAapp study.

## References

1. World Health Organization. Adherence to long-term therapies: evidence for action URL: http://www.who.int/chp/knowledge/publications/adherence_report/en/ [accessed 2019-07-22] [WebCite Cache ID 72wOf9vwx]
2. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
3. van Mierlo T, Fournier R, Ingham M. Targeting Medication Non-Adherence Behavior in Selected Autoimmune Diseases: A Systematic Approach to Digital Health Program Development. PLoS One 2015;10(6):e0129364 [FREE Full text] [doi: 10.1371/journal.pone.0129364] [Medline: 26107637]
4. Struck R, Baumgarten G, Wittmann M. Cost-efficiency of knowledge creation: randomized controlled trials vs. observational studies. Curr Opin Anaesthesiol 2014 Apr;27(2):190-194. [doi: 10.1097/ACO.0000000000000060] [Medline: 24514037]
5. van Asselt T, Ramaekers B, Corro Ramos I, Joore M, Al M, Lesman-Leegte I, et al. Research Costs Investigated: A Study Into the Budgets of Dutch Publicly Funded Drug-Related Research. Pharmacoeconomics 2018 Jan;36(1):105-113 [FREE Full text] [doi: 10.1007/s40273-017-0572-7] [Medline: 28933003]
6. Newington L, Metcalfe A. Factors influencing recruitment to research: qualitative study of the experiences and perceptions of research teams. BMC Med Res Methodol 2014 Jan 23;14:10 [FREE Full text] [doi: 10.1186/1471-2288-14-10] [Medline: 24456229]
7. Compston A, Coles A. Multiple sclerosis. Lancet 2008 Oct 25;372(9648):1502-1517. [doi: 10.1016/S0140-6736(08)61620-7] [Medline: 18970977]
8. Ivanova JI, Bergman RE, Birnbaum HG, Phillips AL, Stewart M, Meletiche DM. Impact of medication adherence to disease-modifying drugs on severe relapse, and direct and indirect costs among employees with multiple sclerosis in the US. J Med Econ 2012;15(3):601-609. [doi: 10.3111/13696998.2012.667027] [Medline: 22376190]
9. Steinberg SC, Faris RJ, Chang CF, Chan A, Tankersley MA. Impact of adherence to interferons in the treatment of multiple sclerosis: a non-experimental, retrospective, cohort study. Clin Drug Investig 2010;30(2):89-100. [doi: 10.2165/11533330-000000000-00000] [Medline: 20067327]
10. Burks J, Marshall TS, Ye X. Adherence to disease-modifying therapies and its impact on relapse, health resource utilization, and costs among patients with multiple sclerosis. Clinicoecon Outcomes Res 2017;9:251-260 [FREE Full text] [doi: 10.2147/CEOR.S130334] [Medline: 28496344]
11. Tan H, Cai Q, Agarwal S, Stephenson JJ, Kamat S. Impact of adherence to disease-modifying therapies on clinical and economic outcomes among patients with multiple sclerosis. Adv Ther 2011 Jan;28(1):51-61. [doi: 10.1007/s12325-010-0093-7] [Medline: 21153000]
12. Devonshire V, Lapierre Y, Macdonell R, Ramo-Tello C, Patti F, Fontoura P, et al. The Global Adherence Project (GAP): a multicenter observational study on adherence to disease-modifying therapies in patients with relapsing-remitting multiple sclerosis. Eur J Neurol 2011 Jan;18(1):69-77. [doi: 10.1111/j.1468-1331.2010.03110.x] [Medline: 20561039]

13. Halpern R, Agarwal S, Dembek C, Borton L, Lopez-Bresnahan M. Comparison of adherence and persistence among multiple sclerosis patients treated with disease-modifying therapies: a retrospective administrative claims analysis. Patient Prefer Adherence 2011 Jan 20;5:73-84 [FREE Full text] [doi: 10.2147/PPA.S15702] [Medline: 21423591]
14. Bayer. MyBETAapp URL: https://www.betaseron.com/betaconnect-system/mybetaapp/ [accessed 2019-07-22]
15. Kleiter I, Lang M, Jeske J, Norenberg C, Stollfuß B, Schürks M. Adherence, satisfaction and functional health status among patients with multiple sclerosis using the BETACONNECT® autoinjector: a prospective observational cohort study. BMC Neurol 2017 Sep 06;17(1):174 [FREE Full text] [doi: 10.1186/s12883-017-0953-8] [Medline: 28877664]
16. Ziemssen T, Sylvester L, Rametta M, Ross AP. Patient Satisfaction with the New Interferon Beta-1b Autoinjector (BETACONNECT™). Neurol Ther 2015 Dec;4(2):125-136 [FREE Full text] [doi: 10.1007/s40120-015-0036-y] [Medline: 26662362]
17. Limmroth V, Bartzokis I, Bonmann E, Kusel P, Schreiner T, Schürks M. The BETACONNECT™ system: MS therapy goes digital. Neurodegener Dis Manag 2018 Dec;8(6):399-410. [doi: 10.2217/nmt-2018-0030] [Medline: 30278827]
18. ClinicalTrials.gov. 2017. Medication Usage and Patient Reported Outcomes Evaluation Via myBETAapp in Patients With Multiple Sclerosis Treated With Betaferon: a Pilot Study (PROmyBETAapp) URL: https://clinicaltrials.gov/ct2/show/NCT03134573 [accessed 2018-10-05] [WebCite Cache ID 72wcbGQMR]
19. Alsop J, Medin J, Cornelissen C, Vormfelde SV, Ziemssen T. Two studies in one: A propensity-score-matched comparison of fingolimod versus interferons and glatiramer acetate using real-world data from the independent German studies, PANGAEA and PEARL. PLoS One 2017;12(5):e0173353 [FREE Full text] [doi: 10.1371/journal.pone.0173353] [Medline: 28475587]
20. Devonshire VA, Feinstein A, Moriarty P. Adherence to interferon β-1a therapy using an electronic self-injector in multiple sclerosis: a multicentre, single-arm, observational, phase IV study. BMC Res Notes 2016 Mar 08;9:148 [FREE Full text] [doi: 10.1186/s13104-016-1948-z] [Medline: 26951043]
21. Higuera L, Carlin CS, Anderson S. Adherence to Disease-Modifying Therapies for Multiple Sclerosis. J Manag Care Spec Pharm 2016 Dec;22(12):1394-1401 [FREE Full text] [doi: 10.18553/jmcp.2016.22.12.1394] [Medline: 27882830]
22. Munsell M, Frean M, Menzin J, Phillips AL. An evaluation of adherence in patients with multiple sclerosis newly initiating treatment with a self-injectable or an oral disease-modifying drug. Patient Prefer Adherence 2017;11:55-62 [FREE Full text] [doi: 10.2147/PPA.S118107] [Medline: 28115831]
23. Sertkaya A, Birkenbach A, Berlind A, Eyraud J. ASPE. 2014. Examination of clinical trial costs and barriers for drug development URL: https://aspe.hhs.gov/report/examination-clinical-trial-costs-and-barriers-drug-development [accessed 2019-07-22] [WebCite Cache ID 73m9oso9u]
24. Aung T, Haynes R, Barton J, Cox J, Murawska A, Murphy K, ASCEND Study Collaborative Group. Cost-effective recruitment methods for a large randomised trial in people with diabetes: A Study of Cardiovascular Events iN Diabetes (ASCEND). Trials 2016 Dec 13;17(1):286 [FREE Full text] [doi: 10.1186/s13063-016-1354-9] [Medline: 27296091]
25. Bower P, Brueton V, Gamble C, Treweek S, Smith CT, Young B, et al. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. Trials 2014;15:399 [FREE Full text] [doi: 10.1186/1745-6215-15-399] [Medline: 25322807]
26. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]
27. Moorcraft SY, Marriott C, Peckitt C, Cunningham D, Chau I, Starling N, et al. Patients' willingness to participate in clinical trials and their views on aspects of cancer research: results of a prospective patient survey. Trials 2016 Jan 09;17:17 [FREE Full text] [doi: 10.1186/s13063-015-1105-3] [Medline: 26745891]
28. Unger JM, Cook E, Tai E, Bleyer A. The Role of Clinical Trial Participation in Cancer Research: Barriers, Evidence, and Strategies. Am Soc Clin Oncol Educ Book 2016;35:185-198 [FREE Full text] [doi: 10.14694/EDBK_156686] [Medline: 27249699]
29. Sacristán JA, Aguarón A, Avendaño-Solá C, Garrido P, Carrión J, Gutiérrez A, et al. Patient involvement in clinical research: why, when, and how. Patient Prefer Adherence 2016;10:631-640 [FREE Full text] [doi: 10.2147/PPA.S104259] [Medline: 27175063]
30. US Food and Drug Administration. FDA's MyStudies Application (App) URL: https://www.fda.gov/Drugs/ScienceResearch/ucm624785.htm [accessed 2018-12-03] [WebCite Cache ID 74OHil3WU]

## Abbreviations

**DMD:** disease-modifying drug
**MS:** multiple sclerosis

*Edited by CL Parra-Calderón; submitted 12.04.19; peer-reviewed by C Alvarez, K Lee Rice; comments to author 23.05.19; revised version received 04.06.19; accepted 15.06.19; published 29.07.19.*

*Please cite as:*
*Limmroth V, Hechenbichler K, Müller C, Schürks M*
*Assessment of Medication Adherence Using a Medical App Among Patients With Multiple Sclerosis Treated With Interferon Beta-1b: Pilot Digital Observational Study (PROmyBETAapp)*
*J Med Internet Res 2019;21(7):e14373*
*URL: http://www.jmir.org/2019/7/e14373/*
*doi:10.2196/14373*
*PMID:31359863*

Original Paper

# Incorporating Information From Electronic and Social Media Into Psychiatric and Psychotherapeutic Patient Care: Survey Among Clinicians

Katherine W Hobbs[1], BA; Patrick J Monette[1], BA; Praise Owoyemi[1], BA; Courtney Beard[1,2], PhD; Scott L Rauch[1,2], MD; Kerry J Ressler[1,2], MD, PhD; Ipsit V Vahia[1,2], MD

[1]McLean Hospital, Belmont, MA, United States

[2]Harvard Medical School, Cambridge, MA, United States

**Corresponding Author:**
Ipsit V Vahia, MD
McLean Hospital
115 Mill Street
Belmont, MA, 02478
United States
Phone: 1 617 855 3291
Email: ivahia@partners.org

## Abstract

**Background:** Obtaining collateral information from a patient is an essential component of providing effective psychiatric and psychotherapeutic care. Research indicates that patients' social and electronic media contains information relevant to their psychotherapy and clinical care. However, it remains unclear to what degree this content is being actively utilized by clinicians as a part of diagnosis or therapy. Moreover, clinicians' attitudes around this practice have not been well characterized.

**Objective:** This survey aimed to establish the current attitudes and behaviors of outpatient clinicians regarding the incorporation of patients' social and electronic media into psychotherapy.

**Methods:** A Web-based survey was sent to outpatient psychotherapists associated with McLean Hospital in Belmont, Massachusetts. The survey asked clinicians to indicate to what extent and with which patients they reviewed patients' social and electronic media content as part of their clinical practice, as well as their reasons for or against doing so.

**Results:** Of the total 115 respondents, 71 (61.7%) indicated that they had viewed at least one patient's social or electronic media as part of psychotherapy, and 65 of those 71 (92%) endorsed being able to provide more effective treatment as a result of this information. The use of either short message service text messages or email was significantly greater than the use of other electronic media platforms ($\chi^2_1$=24.1, n=115, $P$<.001). Moreover, the analysis of survey responses found patterns of use associated with clinicians' years of experience and patient demographics, including age and primary diagnosis.

**Conclusions:** The incorporation of patients' social and electronic media into therapy is currently common practice among clinicians at a large psychiatric teaching hospital. The results of this survey have informed further questions about whether reviewing patient's media impacts the quality and efficacy of clinical care.

***(J Med Internet Res 2019;21(7):e13218)*** doi:10.2196/13218

**KEYWORDS**

technology; social media; psychotherapy; psychiatry

## Introduction

Electronic and social media platforms have become ubiquitous and essential tools in navigating the 21st century [1]. Content stored and shared on these platforms, therefore, contains extensive information about our everyday lives and interactions [2]. Research in the mental health field has begun to explore whether electronic and social media content may contain clinically relevant markers of individuals' behavior and mental health. During this period of rapid uptake in utilization of these platforms, it is important to understand how they may reflect, impact, or be used to augment treatment for mental health conditions. In this study, we use the term electronic and social media platforms to refer to apps that individuals access on their mobile phones, tablets, or computers for the purpose of communicating with or sharing content with others. The most prominent examples of these platforms include Facebook, Instagram, email, short message service (SMS) text messaging, and other messaging apps.

A growing body of literature has begun to examine how we may leverage electronic media usage to identify markers of psychiatric illness or response to treatment. Studies analyzing electronic media content have identified differences in language use between healthy individuals and individuals with different psychiatric disorders, suggesting that clinically relevant signals across a wide variety of illnesses can be found in the language used on social media platforms [3,4]. Likewise, a recent study found that analysis of language in Facebook posts can predict the existence of a depression diagnosis in medical records with a similar degree of accuracy as established depression screening surveys [5]. Another study noted that patients' word usage in emails was a predictor of therapeutic outcomes in a sample of individuals with social anxiety [6]. Another study correlated the frequency of social media posting with health outcomes—those who posted most frequently on Facebook in their sample were more likely to have a diagnosis or positively screen for depression [7].

In examining social media postings, researchers have developed computational models to better predict onsets of psychiatric illnesses. One such study analyzed the Twitter activity of users who reported a clinical depression diagnosis [8]. Researchers retroactively examined participants' Twitter content from the year before the onset of depression to measure markers of social activity, emotions, and relational concerns. They utilized these markers to create a model predicting the likelihood of depression onset before a clinical diagnosis [8]. In a similar study, researchers created computational methods to screen for markers of depression and posttraumatic stress disorder (PTSD) in Twitter users' posts [9]. They reviewed the Twitter data of individuals with depression and PTSD from the year before their clinical diagnosis and those of healthy subjects. They found their computational models could distinguish content from healthy and depressed subjects and demonstrated improved accuracy in correctly diagnosing depression and PTSD as compared with the general performance of practitioners [9].

Simultaneously, research has begun to explore patient attitudes toward sharing these electronic media data in the context of clinical care. A recent study surveyed new patients in a psychiatric outpatient setting regarding their use of mobile phones and their willingness to share data collected on their phones with mental health providers [10]. Overall, the survey found moderate willingness among patients in a psychiatric outpatient setting to share data collected via a mobile phone with their clinicians and, unsurprisingly, respondents were more likely to grant a mental health app access to less personal content [10]. Research has also explored mental health providers' general attitudes toward gathering information about their patients through electronic sources but has not explicitly addressed how clinicians are utilizing patients' own social media content to inform care [11]. For example, in one study, 20% of responding psychiatrists and psychologists reported searching online for information about their patients either *sometimes* or *often* [11]. Of these clinicians, 45% said they would perform a search to cross-check information, 35% would do so out of curiosity, and 60% would do so to gather information. The study did not specify what type of information clinicians were gathering, nor how this information impacted the care they delivered [11].

Studies such as these indicate that electronic and social media can be leveraged as effective predictors and trackers of mental illness, but it remains unclear to what degree this type of content is being actively utilized by clinicians as a part of diagnosis or therapy. Thus, gaps remain regarding how mental health clinicians are utilizing patients' electronic and social media to identify clinically relevant information to improve care and how they are weighing potential benefits and drawbacks of doing so. To address these gaps, we surveyed mental health clinicians who provide outpatient psychotherapy. The goal of the survey was to establish to what degree clinicians are currently accessing patients' social and electronic media content, how they are doing so, in what ways they are incorporating this content into therapy, and what concerns they have around this practice. Notably, our project did not seek to characterize electronic communication between clinicians and patients for the purpose of scheduling, follow-up, and other clinical needs.

## Methods

### Respondents

The survey population consisted of 115 outpatient psychotherapy clinicians associated with McLean Hospital in Belmont, Massachusetts, who completed a Web-based survey regarding their use of social and electronic media as part of their standard clinical care. The principal investigator emailed the survey to the 244 individuals included on the McLean Hospital master outpatient clinician mailing list. Emails were sent out in 2 rounds of email blasts between April 19, 2018, and June 6, 2018. The first round of emails went to 102 outpatient clinicians who work within formal hospital clinics, and the second went to 142 clinicians doing hospital-based private practice. For both rounds, reminder emails were sent 3, 5, and 8 days following the original email. Of the 244 email recipients, 13 email addresses were deemed undeliverable, resulting in 231 valid email addresses. With 115 total respondents out of 231 recipients, the overall response rate for this survey was 49.8%. The respondents spanned a range of professional disciplines and reflected the diversity of psychotherapy provided at McLean hospital. The email link to the survey specified that the purpose of the survey was to gain an understanding of whether and how clinicians were accessing electronic media in regular care. We did not collect demographic information on clinicians.

### Survey Instrument

We were unable to identify an existing instrument that would adequately allow us to capture the breadth of information regarding clinician attitudes and behaviors around use of electronic media. Hence, we developed our own instrument designed by consensus among senior study investigators (KR, CB, and IV) to capture naturalistic use of electronic media in therapy. We built in items to facilitate stratification by qualification and years of experience. We elected to limit information about clinician demographics to minimize survey time and maximize response rate. We next built items to capture what types of electronic media platforms were used and in which patient populations this approach had been tried. Finally, we developed items to capture with greater detail the process of accessing this information, maintaining privacy, and its usefulness.

The survey, which is available in Multimedia Appendix 1, was created using Google Forms and consisted of 17 questions. First, respondents were asked to indicate their professional degree and number of years of clinical experience. Respondents were then asked to respond yes or no to whether they had viewed a patient's social or electronic media as part of outpatient psychotherapy. If they responded no, they were asked whether or not they had considered viewing this content and the major factors or concerns they might consider in making the decision.

If they responded yes, they were prompted to answer 12 more questions aimed at characterizing their use of this content in psychotherapy. Clinicians were asked which platforms they accessed and how they accessed them—whether it was directly with the patient in the session, outside of the session with the patient's permission, or indirectly via report from the patient or their loved ones. The survey also asked with which age demographic and clinical populations the clinicians used this approach, with how many patients they used it, and in approximately how many sessions per patient. Following these questions, clinicians were asked whether they believed access to this information helped improve the quality of care they delivered. Additional details were collected around the process of accessing this information—namely, which party had suggested access to this content and whether they had discussed issues around privacy. Finally, 2 free-response questions allowed respondents to share their reasons for and concerns about accessing this type of content in their clinical care.

### Data Safety and Storage

The survey was distributed via a private Google Survey invitation to clinicians' professional secure email addresses. Only those with the link were able to access it. Data from the survey were downloaded and saved on private Partners Healthcare servers. No identifying information was collected, unless the respondents chose to volunteer such information at the end of the survey; there was space for respondents to give their name and email address if they were interested in learning more about technology-related projects being conducted at the hospital. This question was removed from the dataset before data analysis to deidentify the responses.

### Data Analysis

All data analyses were performed using the SPSS 3 (IBM SPSS Statistics) statistical package. For quantitative items, descriptive statistics were generated, and then respondents were stratified by years of clinical experience and by professional degree, followed by a comparison of differences using chi-square tests. For questions in which respondents had the option to write free-text responses, these responses were consolidated into categories upon consensus by study staff and then recoded accordingly. As the volume of qualitative data was relatively low and we had distinct statements from each respondent, we elected not to use customized qualitative analysis software. Instead, we adapted a simple qualitative classification approach [12] and sorted qualitative statements into broad themes which this study investigators agreed to by consensus.

This project was undertaken as a quality improvement initiative at McLean Hospital and as such was not formally supervised by the institutional review board per their policies. The institutional review board provides a checklist [13] to determine whether a specific project qualifies as a quality improvement initiative. A total of 4 of the study investigators (KH, PM, PO, and IVV) assessed this study simultaneously and independently using this checklist, and all determined that the project met criteria to qualify as a QII. We also communicated this to the institutional review board who confirmed that if the study meant QII, formal supervision was not required. We utilized the existing mechanism for mass email communication with McLean clinicians to appropriately disseminate the survey, and we specified that responding to the survey was voluntary and part of an information gathering process.

## Results

### Respondent Demographics

Of the 115 clinicians who responded to the survey, 31 (27.0%) hold MDs as their professional degree, 35 (30.4%) hold PhDs, and 30 (26.1%) hold Licensed Clinical Social Worker (LCSW) degrees. Other degrees represented in smaller numbers include: Advanced Practice Registered Nursing degrees, MD/PhD, Licensed Mental Health Counselor degrees, and MA degrees. Moreover, 47 (40.9%) respondents have less than 10 years of clinical experience, 29 (25.2%) have 10 to 20 years of experience, and 39 (33.9%) have more than 20 years of experience.

### Overall Viewing Rates

Of the 115 respondents, 71 (61.7%) indicated that they had viewed at least one patient's social or electronic media as part of psychotherapy. The remaining 44 (38.3%) indicated that they had never viewed a patient's electronic media. A total of 1 respondent who indicated having viewed a patient's media did not answer all the subsequent questions regarding this content. Therefore, some of the follow-up questions for clinicians who reported yes to viewing media had 71 total responses and some had only 70 responses.

Of the 44 respondents who have not viewed patient's media, 10 (23%) said that they have considered incorporating this content into therapy sessions.

### Types of Media Viewed

Table 1 summarizes the breakdown of patients' social and electronic media platforms that clinicians viewed. A total of 60 of 70 respondents to this question (86%) viewed SMS text messages and 56 (80%) viewed emails, making these the most frequently accessed platforms. The use of either SMS text messages or email was significantly greater than the use of any of the other electronic media platforms ($\chi^2_1$=24.1, N=115, *P*<.001).

**Table 1.** Type of media viewed by clinicians.

| Media type | Responses (yes), n (%) |
|---|---|
| SMS (short service message) | 60 (86) |
| Email | 56 (80) |
| Facebook | 27 (39) |
| Call history | 13 (19) |
| Instagram | 12 (17) |
| Blogs | 7 (10) |
| Twitter | 7 (10) |
| Snapchat | 6 (9) |
| WhatsApp | 5 (7) |

## Clinical Experience and Media Use

A summary of media viewing according to respondents' clinical experience is shown in Table 2. A total of 33 of the 47 respondents (70%) with less than 10 years of experience and 20 of the 29 respondents (69%) with 10 to 20 years of experience stated they had viewed a patient's social or electronic media; meanwhile, 18 of the 39 respondents (46%) with more than 20 years of experience reported having done so. The relationship between years of experience and viewing rates was significant ($\chi^2_2$=6.1, N=115, *P*=.048). We also performed similar analyses stratifying clinicians by their professional degree but did not note any significant differences between groups or even nonstatistically significant trends.

## Methods of Viewing Media

Of the 70 respondents to this question who have viewed patients' electronic or social media, 62 (89%) reported having viewed this content directly in the patients' presence, and 46 (66%) indicated having gathered information about the media content through patient self-report, as demonstrated in Table 3.

**Table 2.** Overall clinician utilization of patient electronic media information in psychotherapy.

| Media type[a] | Total yes responses, n (%) | <10 years clinical experience, n (%) | 10-20 years clinical experience, n (%) | 20+ years clinical experience, n (%) | Chi-square (df) | *P* value (2-sided) |
|---|---|---|---|---|---|---|
| Email (n=70) | 56 (80) | 23 (33) | 17 (24) | 16 (23) | 2.5 (2) | .28 |
| Text (n=70) | 60 (86) | 28 (40) | 17 (24) | 15 (21) | 0.2 (2) | .92 |
| Facebook (n=70) | 27 (39) | 14 (20) | 8 (11) | 5 (7) | 1.3 (2) | .53 |
| Instagram(n=70) | 12 (17) | 5 (7) | 4 (6) | 3 (4) | 0.2 (2) | .92 |
| Call history (n=70) | 13 (19) | 4 (6) | 3 (4) | 6 (9) | 3.5 (2) | .17 |
| Whatsapp (n=70) | 5 (7) | 1 (1) | 3 (4) | 1 (1) | 2.7 (2) | .26 |
| Twitter (n=70) | 7 (10) | 4 (6) | 1 (1) | 2 (3) | 0.8 (2) | .67 |
| Blogs (n=70) | 7 (10) | 0 (0) | 3 (4) | 4 (6) | 7.1 (2) | .03 |
| Snapchat (n=70) | 6 (9) | 2 (3) | 1 (1) | 3 (4) | 2.1 (2) | .36 |
| Any | 70 (100) | 32 (46) | 20 (29) | 18 (25) | 5.4 (2) | .07 |

[a]We used chi-square tests to compare clinicians who responded *yes* with using a specific media platform by years of clinical experience.

**Table 3.** Clinician's methods of accessing patients' electronic or social media content.

| Method of access | Responses (yes), n (%) |
|---|---|
| Viewed content directly in patient's presence | 62 (89) |
| Patient self-report | 46 (66) |
| Outside of session with permission | 14 (20) |
| Report from friends or relatives | 12 (17) |

### Patient Demographic

Clinicians reported viewing social or electronic media with the following patient age groups: 20 of 70 (29%) adolescents, 46 (66%) young adults, 46 (66%) adults, and 5 (7%) older adults.

Moreover, clinicians indicated accessing this content most commonly with patients diagnosed with depression or anxiety; 42 of 70 (60%) of yes respondents stated that they viewed media with patients diagnosed with either or both conditions, followed by 32 (46%) with borderline personality disorder. Each of the remaining diagnoses yielded less than 35% of yes responses, as indicated in Tables 4 and 5.

### Viewing Frequency

A total of 9 of 70 (13%) clinicians who viewed their patient's electronic or social media reported doing so only 1 time, whereas 40 (57%) clinicians reported doing so very infrequently. A total of 18 of 70 (26%) reported viewing media content in roughly every 5 to 10 sessions per patient, and 2 (3%) reported viewing it every 2 to 3 sessions. Finally, 1 (1%) reported viewing this content in every session.

Moreover, 12 of 70 (17%) clinicians who viewed their patient's electronic or social media reported doing so with just 1 to 2 patients, 26 (37%) reported doing so with 3 to 5 patients, 14 (20%) reported doing so with 6 to 10 patients, and 18 (26%) reported doing so with more than 10 patients.

### Whose Idea?

A total of 52 of 70 (74%) clinicians reported that it was the patient's idea to incorporate social and electronic media into psychotherapy, whereas 10 (14%) clinicians reported that it was their own idea. A total of 7 (10%) indicated that it was a mutual suggestion, 1 (1%) reported that it was a family member's idea, and 1 (1%) indicated that multiple parties suggested it.

### Impact on Treatment

We incorporated a single item asking clinicians to rate the extent to which they were able to provide more effective treatment in part because of accessing their patients' electronic or social media. A total of 17 of 70 (24%) clinicians reported noting *significant* improvement in the level of care, whereas 30 (42%) reported seeing *moderate* improvement, 18 (25%) reported *slight* improvement, and 6 (8%) reported *no improvement* in their ability to deliver effective care. There were no differences on this measure when we stratified clinicians by years of experience or highest level of training.

### Privacy

We inquired whether clinicians had discussed privacy with patients and, if so, who initiated the conversation. A total of 44 of 71 (62%) clinicians who had accessed patients' social and electronic media indicated that they discussed issues of privacy with the patients. We found that clinicians with greater experience were significantly more likely ($\chi^2_4$=13.2, N=115, $P$=.01) to bring up privacy concerns with their patients; 16 of the 18 (89%) clinicians with more than 20 years of experience who viewed this content reported having a conversation around privacy, whereas 12 of 20 (60%) of those with between 10 and 20 years of experience and 16 of 33 (48%) of those with less than 10 years of experience who viewed this content reported doing so.

The majority of clinicians who discussed privacy initiated the conversations themselves; only 2 of 71 (3%) clinicians reported that their patients raised privacy concerns regarding sharing this type of content.

### Reasons for Accessing This Content

Of the 71 respondents who indicated having accessed this content, 63 (89%) provided free-text explanations of their motivations for accessing patients' media as part of clinical care. Using a qualitative analytic approach described in the Methods section, we categorized those responses by consensus into the following 5 general thematic categories, listed below in order of frequency. If comments fell into multiple categories, they were counted under each relevant category. Therefore, the percentages add to 117, rather than 100.

1. To monitor or address a specific target behavior (26/63, 41%)
2. To provide feedback on patients' electronic communications and behaviors (26/63, 41%)
3. To obtain collateral information on patients' life (15/63, 24%)
4. To establish working alliance/rapport (4/63, 6%)
5. Logistical reasons (3/63, 5%)

### Concerns About Accessing This Content

Of the 71 respondents who indicated having accessed this content, 48 (68%) provided free-text explanations of their concerns regarding accessing this content. On the basis of our qualitative analytic approach described in the Methods section, responses fell into 7 general thematic categories, listed below in order of frequency. In this case, all responses fell into just 1 category, and thus, were counted once.

1. Privacy concerns/ethical boundary (15/48, 31%)
2. No concerns as long as done on patients' own terms (10/48, 21%)
3. Detracts or distracts from therapy (7/48, 15%)
4. Content is subjective and easily misinterpreted (6/48, 13%)
5. None (6/48, 13%)
6. Time constraint (2/48, 4%)
7. Other (2/48, 4%)

**Table 4.** Media viewing by patient age demographic

| Patient age demographic | Responses (yes), n (%) |
|---|---|
| Adults | 46 (66) |
| Young adults | 46 (66) |
| Adolescents | 20 (29) |
| Older adults | 5 (7) |

**Table 5.** Media viewing by patient diagnosis.

| Patient diagnosis | Responses (yes), n (%) |
|---|---|
| Anxiety | 42 (60) |
| Depression | 42 (60) |
| Borderline personality disorder | 32 (46) |
| Posttraumatic stress disorder | 22 (31) |
| Bipolar disorder | 21 (30) |
| Psychotic disorders | 17 (24) |
| Obsessive compulsive disorder | 42 (60) |
| Eating disorders | 7 (10) |

## *Discussion*

### Conclusions

This survey addresses the naturalistic use of patients' electronic and social media by mental health clinicians. With a 49.8% (115/231) response rate, the survey results provide useful insights into the current practices among a diverse group of therapists. We noted that the majority of outpatient clinicians surveyed (115/71, 61.7%) reported having viewed at least one patient's electronic or social media as part of care, with email and SMS text messaging emerging as the most frequently accessed platforms by far. We also found that clinicians who had been in practice for fewer years accessed this information more frequently than more experienced clinicians, recognizing that this is confounded by age cohort of the clinicians as well. This may reflect a greater proficiency among more junior clinicians in understanding how technology may be intrinsic to daily life [14,15]. Conversely, this may also reflect a lesser-perceived need for concrete collateral information among more experienced therapists, who may practice with a better-established frame [16].

Another prominent finding was that only 3% of clinicians who reported viewing social or electronic media indicated that any of their patients had voiced concerns around privacy. This was a surprisingly low number, which may reflect patient confidence in allowing therapists to access this information, or a hesitancy to disagree with a therapist's suggestion because of power dynamics. It may also be related to our finding that the majority of clinicians accessed these data during sessions, in the patients' presence. Accessing such personal information face to-face and giving patients control over the ability to share this information may have helped foster a sense of the privacy of the session extending to electronic data as well [17]. The ability to access collateral information, whether from family members [18] or through patient writings or other forms of expressive therapy [19], has long been a tradition in psychotherapy. Accessing electronic communications represents an extension of this tradition, incorporating contemporary means of communication and leveraging technology that may enable extraction on deeper behavioral signals from natural language.

Overall, our finding that a high percentage of clinicians are accessing electronic media and doing so with patients present—and in over 74% of cases, at patients' suggestion—indicates that the process of incorporating this information into therapy is a common and organic process across various provider types (MD, PhD, LCSW). Clinicians reported accessing more private data, such as email and SMS text messaging, to a greater extent than social media, which may indicate clinicians' perceptions that these more personal platforms contain more relevant behavioral signals. Our finding that 92% of clinicians reported that accessing electronic or social media improved their ability to provide effective treatment points to a role for increased attention to electronic media in care. Therefore, there is an imperative to explore whether this perception of benefit holds true in observational and experimental studies, which quantify potential benefit in a controlled manner and to understand the mechanisms underlying any observed benefit. Future studies are also needed to establish best practice standards that guide the appropriate access and use of this type of content across different providers.

### Limitations

Our findings should be interpreted within the context of the limitations of this survey. Although our respondents spanned a range of degrees and years of experience, we only surveyed clinicians affiliated with a single psychiatric institution in the Boston region, which may impact the generalizability of the findings. Furthermore, the patient population served by McLean Hospital therapists may not be representative of the population at large, which in turn may impact the representativeness of the sample with regard to use of electronic media as well.

In addition, our survey was designed to be brief, anonymous, and easy to complete to maximize response rates, rather than to be comprehensive. Consequently, we did not obtain details on provider demographics. We also did not ask clinicians to share the breakdown of diagnoses or age groups that they see in their outpatient practices, so results regarding media viewing by these variables may be skewed by the makeup of the overall patient population seen by clinicians in our sample. Collection of these data would be important in follow-up work. Although we do not believe this is a common occurrence, we did not specifically inquire whether clinicians accessed patients' data without patients' permission.

We acknowledge that our survey did not address all of the ways in which clinicians may be utilizing electronic platforms to obtain clinical information about a patient's wellbeing. Specifically, we did not ask about the use of apps to track patients' mood and behavior over time, for example, through daily mood or activity logs on one's mobile phone. Apps of this type are emerging within psychotherapy, and clinicians may utilize these behavioral tracking methods in addition to or in lieu of the informal methods of accessing and discussing patients' social and electronic media content that we laid out in the survey.

Finally, limitations associated with surveys, including responder bias and recall bias, apply to this work as well. As we did not track provider demographics, and the survey was anonymized, we are not able to compare responders with nonresponders and draw insights into the effects of responder bias. We recommend that future work, including from our team, study this issue to gain a better understanding of predictors of using electronic media in therapy.

Despite these limitations, our work provides an early indication that therapists in practice are incorporating information from digital platforms into the care process. Although evidence regarding the impact of this information is limited, our findings identified 2 primary purposes for accessing the information—to monitor and address a specific target behavior and to provide feedback on patients' electronic communications and behaviors. This finding suggests that the additional information serves as a valuable tool in the therapeutic process beyond just collecting collateral information.

### Future Directions

Future work on this topic should focus on replicating our initial findings in larger, more diverse clinician populations. More data are also needed to determine, in a more specific manner, exactly how and when electronic media may enhance or hinder the therapy process. Issues around privacy and confidentiality of this information also merit thoughtful discussion. This, in turn, may help develop a more systematic approach toward optimally utilizing electronic and social media in the augmentation of the therapeutic process in mental health care settings.

## Acknowledgments

The authors would like to thank all the clinicians who took the time to complete their survey, provide feedback, and support their research. They appreciate input from Benjamin Silverman, MD, who advised their team on this survey's exemption from institutional review board approval.

## Conflicts of Interest

IVV receives research funding from the Once Upon a Time Foundation and the Massachusetts Institute of Technology. He also receives honorarium for his editorial role on the American Journal of Geriatric Psychiatry. KJR is on the scientific advisory boards for the Sheppard Pratt-Lieber Research Institute, the Laureate Institute for Brain Research, the Army Study to Assess Risk and Resilience in Servicemembers project, the University of California–San Diego VA Center of Excellence for Stress and Mental Health, and the Anxiety and Depression Association of America. He provides fee-for-service consultation for Biogen and Alkermes. SLR is principally employed by McLean Hospital/Partners Healthcare. He receives a stipend from Society of Biological Psychiatry for serving as secretary, receives royalties from Oxford University Press and American Psychiatric Publishing Inc, provides paid service as a member of a Research Advisory Committee for the VA, provides unpaid service on the governing boards of the National Network of Depression Centers and the Anxiety and Depression Association of America, and conducts research supported by the National Institutes of Health. KH, PO, PM, and CB each declare no potential conflicts of interest.

## Multimedia Appendix 1

Text of the survey that was sent to clinicians via a Google Form.

[PDF File (Adobe PDF File), 61KB - jmir_v21i7e13218_app1.pdf ]

## References

1. Chaffrey D. Smart Insights. 2019. Global Social Media Research Summary 2019 URL: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

2. Gross R, Acquisti A. Information Revelation and Privacy in Online Social Networks. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society. 2005 Presented at: WPES'05; November 7, 2005; Alexandria, VA, USA p. 71-80. [doi: 10.1145/1102199.1102214]
3. Coppersmith G, Dredze M, Harman C. Quantifying Mental Health Signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: CLPSYCH'14; June 27, 2014; Baltimore, Maryland, USA p. 51-60. [doi: 10.3115/v1/W14-3207]
4. Guntuku S, Ramsay J, Merchant R, Ungar LH. Language of ADHD in adults on social media. J Atten Disord 2017 Nov 1:1087054717738083 (forthcoming). [doi: 10.1177/1087054717738083] [Medline: 29115168]
5. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci USA 2018 Dec 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
6. Hoogendoorn M, Berger T, Schulz A, Stolz T, Szolovits P. Predicting social anxiety treatment outcome based on therapeutic email conversations. IEEE J Biomed Health Inform 2017 Dec;21(5):1449-1459 [FREE Full text] [doi: 10.1109/JBHI.2016.2601123] [Medline: 27542187]
7. Smith RJ, Crutchley P, Schwartz HA, Ungar L, Shofer F, Padrez KA, et al. Variations in Facebook posting patterns across validated patient health conditions: a prospective cohort study. J Med Internet Res 2017 Dec 6;19(1):e7 [FREE Full text] [doi: 10.2196/jmir.6486] [Medline: 28062392]
8. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media. 2013 Presented at: ICWSM'13; July 8–11, 2013; Cambridge, Massachusetts, USA.
9. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Dec 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
10. di Matteo D, Fine A, Fotinos K, Rose J, Katzman M. Patient willingness to consent to mobile phone data collection for mental health apps: structured questionnaire. JMIR Ment Health 2018 Aug 29;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]
11. Deen SR, Withers A, Hellerstein DJ. Mental health practitioners' use and attitudes regarding the internet and social media. J Psychiatr Pract 2013 Nov;19(6):454-463. [doi: 10.1097/01.pra.0000438184.74359.88] [Medline: 24241499]
12. Jeste DV, Vahia IV. Comparison of the conceptualization of wisdom in ancient Indian literature with modern views: focus on the Bhagavad Gita. Psychiatry 2008;71(3):197-209 [FREE Full text] [doi: 10.1521/psyc.2008.71.3.197] [Medline: 18834271]
13. Hohmann EL. Partners HealthCare. 2012. Clinical Quality Improvement/Measurement Checklist URL: https://www.partners.org/Assets/Documents/Medical-Research/Clinical-Research/Clinical-QI-Checklist.doc
14. Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 Oct;78(10):645-655. [doi: 10.1016/j.ijmedinf.2009.04.008] [Medline: 19501017]
15. Jent JF, Eaton CK, Merrick MT, Englebert NE, Dandes SK, Chapman AV, et al. The decision to access patient information from a social media site: what would you do? J Adolesc Health 2011 Oct;49(4):414-420 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.02.004] [Medline: 21939873]
16. Masters K. For what purpose and reasons do doctors use the internet: a systematic review. Int J Med Inform 2008 Jan;77(1):4-16. [doi: 10.1016/j.ijmedinf.2006.10.002] [Medline: 17137833]
17. Lehavot K, Barnett JE, Powers D. Psychotherapy, professional relationships, and ethical considerations in the MySpace generation. Prof Psychol Res Pr 2010;41(2):160-166. [doi: 10.1037/a0018709]
18. Hankoff LD, Engelhardt DM, Freedman N, Mann D, Margolis R. Denial of illness in schizophrenic out-patients: effects of psychopharmacological treatment. Arch Gen Psychiatry 1960 Dec;3:657-664. [doi: 10.1001/archpsyc.1960.01710060089013] [Medline: 13711239]
19. Lawver T. A proposal for including patient-generated web-based creative writing material into psychotherapy: advantages and challenges. Psychiatry (Edgmont) 2008 Jun;5(6):56-61 [FREE Full text] [Medline: 19727286]


## Abbreviations

**LCSW:** Licensed Clinical Social Worker
**PTSD:** posttraumatic stress disorder
**SMS:** short message service

*Edited by G Eysenbach; submitted 21.12.18; peer-reviewed by K Metsaranta, D Di Matteo; comments to author 08.04.19; revised version received 26.04.19; accepted 17.05.19; published 12.07.19.*

*Please cite as:*
*Hobbs KW, Monette PJ, Owoyemi P, Beard C, Rauch SL, Ressler KJ, Vahia IV*
*Incorporating Information From Electronic and Social Media Into Psychiatric and Psychotherapeutic Patient Care: Survey Among Clinicians*
*J Med Internet Res 2019;21(7):e13218*
*URL: http://www.jmir.org/2019/7/e13218/*
*doi:10.2196/13218*
*PMID:31301127*

Original Paper

# Mothers' Perceptions of the Internet and Social Media as Sources of Parenting and Health Information: Qualitative Study

Rachel Y Moon[1], MD; Anita Mathews[2], MS; Rosalind Oden[2]; Rebecca Carlin[2], MD

[1]University of Virginia, Charlottesville, VA, United States

[2]Children's National Health System, Washington, DC, United States

**Corresponding Author:**
Rachel Y Moon, MD
University of Virginia
PO Box 800386
Charlottesville, VA, 22908
United States
Phone: 1 4349245521
Email: rym4z@virginia.edu

## Abstract

**Background:** Traditionally, guidance and support to new parents have come from family, friends, and health care providers. However, the internet and social media are growing sources of guidance and support for parents. Little is known about how the internet and social media are used by parents of young infants and specifically about parental perceptions of the internet and social media as sources of parenting and infant health information.

**Objective:** The aim of this study was to explore, using qualitative methods, parental perceptions of the advantages and disadvantages of the internet and social media as sources of parenting and health information regarding their infant.

**Methods:** A total of 28 mothers participated in focus groups or individual interviews. Probing questions concerning parenting and health information sources were asked. Themes were developed in an iterative manner from coded data.

**Results:** The central themes were (1) reasons that mothers turn to the internet for parenting and health information, (2) cautionary advice about the internet, and (3) reasons that mothers turn to social media for parenting and health information. Mothers appreciated the ability to gather unlimited information and multiple opinions quickly and anonymously, but recognized the need to use reputable sources of information. Mothers also appreciated the immediacy of affirmation, support, and tailored information available through social media.

**Conclusions:** The internet and social media are rapidly becoming important and trusted sources of parenting and health information that mothers turn to when making infant care decisions.

***(J Med Internet Res 2019;21(7):e14289)*** doi:10.2196/14289

**KEYWORDS**

internet; parenting; social media; focus groups

## Introduction

Most adults find parenting to be an entirely new experience for which they have little background knowledge. It can easily become overwhelming, and parents need guidance and support. Traditionally, guidance and support have come from family, friends, and health care providers [1-4]. However, just as the internet is becoming a growing source of health information in general, with 59% of US adults seeking health information on the internet [5], it is also a growing source of guidance and support for parents [6,7].

In addition, the recent advent of social media (defined as forms of electronic communication through which one can share information, ideas, and personal messages with others online [8]), including online social networks (eg, Facebook), email listservs, blogs, and mobile phone apps, has introduced new channels through which one can seek information and the opinions of others. Parents, in particular, have found these interactive forums helpful. The majority of US parents who use social media state that it provides useful parenting information, and almost half have received support on social media regarding a parenting issue in the past month [9]. Mothers of young infants find social networking sites to be important sources of social support [10].

Our previous research has found that mothers are more consistently the primary decision maker pertaining to the infant [1]. In addition, women are more likely to seek advice and help from multiple sources, including social media [9], whereas men are more likely to depend almost solely upon their spouses [11,12]. Little is known about how the internet and social media are used by mothers of young infants as sources of parenting and health information. Thus, we conducted a qualitative study to explore maternal perceptions of the internet and social media as parenting and health information sources.

## Methods

### Recruitment

Mothers of healthy term infants <6 months of age were recruited to participate in a larger quantitative survey about their personal social networks if they were English speaking, African American (AA) or Caucasian, aged >18 years, a primary caregiver for the infant, and lived in the metropolitan Washington DC area. Each mother signed a written informed consent upon enrollment. From this sample, we selected a subsample [13] to participate in focus groups or individual interviews. We purposefully chose both AA and Caucasian mothers with different educational and socioeconomic levels to assure a wide range of attitudes and opinions. Socioeconomic levels were dichotomized (lower and higher) based on enrollment in public health insurance (Medicaid or the equivalent) and eligibility for the Special Supplemental Nutrition Program for Women, Infants, and Children. These proxies were used because eligibility for both programs are income based and because eligibility is easily verifiable.

#### *Data Collection*

We conducted all interviews between July 2016 and January 2018. We conducted both focus group interviews and individual in-depth semistructured interviews to accommodate mothers' schedules. Mothers who could not participate in focus groups were offered individual interviews. Focus groups were stratified by race and parity (primiparous or multiparous) to maximize homogeneity of each group's participants, as this can result in increased willingness to share thoughts and opinions [14]. Interview questions were formulated by all authors in group meetings; the same interview guide was used for both interview formats. Questions were modified iteratively based on data collected in previous interviews. Trained facilitators (RO and AM) conducted all the interviews, asking broad, open-ended questions (eg, "What sources, other than your family and friends, do you use for advice?") followed by more specific, probing questions (eg, "Why that source, in particular?") to clarify responses. Focus group and individual interviews averaged 2 hours and 90 min, respectively, in duration. Each participant received a US $75 gift card for their time. This study was approved by the institutional review boards of Children's National Medical Center and the University of Virginia.

### Data Analysis

All interviews were video- and audio-recorded and transcribed by a Health Insurance Portability and Accountability Act–compliant transcription company, after which one author (RO) simultaneously reviewed the video- and audio-recordings and transcript of each interview for accuracy. Any disagreements about the transcription were resolved by consensus after additional authors listened to the recordings. This multistep process was used to maximize accuracy and eliminate bias from the transcription process. Using standard qualitative analytic techniques and a grounded theory approach, transcripts were analyzed line-by-line by the 4 authors, all of whom have previous experience with [1,15-18] or training in qualitative analysis. Qualitative analysis software (NVivo 11 plus [19]) was used to organize, sort, and code the data (quotations). Themes were developed and revised iteratively, as patterns within data emerged [20]. Authors, in regular meetings, discussed emerging themes and patterns in the data and reached a consensus on the major themes. Individual and focus group interviews were analyzed separately, and emerging themes were compared. To increase rigor, concurrent triangulation or use of multiple sources for verification of findings [21] of the focus group interviews and the individual interviews was used to corroborate findings [22]. In addition, we confirmed the findings by peer review and feedback during presentations to child health professionals, pediatric researchers, and community members.

## Results

### Demographics

A total of 8 focus groups and 2 individual interviews with 28 mothers (26 participated in focus groups [median 3.5 participants, range 2-6 participants]; 2 participated in individual interviews) were conducted, and thematic saturation was reached. At the time of the interview, mothers had a mean age of 30.4 years (range 20-44 years), 71.4% of the mothers were AA (which reflects the racial distribution in the larger sample and is representative of the general population in Washington DC), and slightly over half of the mothers were primiparous; 61% of women were married, and 64.3% of infants were male (see Table 1).

**Table 1.** Characteristics of participants (N=28).

| Characteristics | Statistics |
|---|---|
| Age (years), mean (range) | 30.4 (20-44) |
| **Age (years), n (%)** | |
| 18-24 | 5 (18) |
| 25-29 | 10 (36) |
| 30-34 | 5 (18) |
| ≥35 | 8 (29) |
| **Race or ethnicity, n (%)** | |
| Black or African American | 20 (71) |
| Caucasian | 8 (29) |
| **Educational level, n (%)** | |
| Did not complete high school | 1 (4) |
| High school graduate | 6 (21) |
| Some college | 4 (14) |
| 4-year college graduate | 17 (61) |
| **Socioeconomic status, n (%)** | |
| Lower | 10 (36) |
| Higher | 18 (64) |
| **Marital status, n (%)** | |
| Married | 17 (61) |
| Never married | 10 (36) |
| Separated/divorced | 1 (4) |
| **Number of children, n (%)** | |
| 1 | 15 (54) |
| 2 or more | 13 (46) |
| **Infant gender, n (%)** | |
| Male | 18 (64) |
| Female | 10 (36) |
| Infant age (months), mean (range) | 4.8 (1-11) |
| **Infant age (months), n (%)** | |
| <3 | 7 (25) |
| 3-6 | 15 (54) |
| >6 | 5 (18) |

## Central Themes

Several main themes and subthemes with regard to maternal use of the internet and social media are listed in Table 2. The central themes were (1) reasons that mothers turn to the internet for parenting and health information, (2) cautionary advice about the internet, and (3) reasons that mothers turn to social media for parenting and health information. These themes, with illustrative quotes, are discussed below.

### Reasons That Mothers Turn to the Internet for Parenting and Health Information for Their Infant

Mothers reported that they often had multiple questions about their infant every day, particularly when they were new at mothering, and they appreciated that there was unlimited information at their fingertips:

> *[The Internet] is also good for informational purposes...there's nothing wrong with having an endless amount of information at your disposal. What's wrong with educating yourself on these sorts of things? And so, I believe in higher learning so I*

*can have information at any way possible. And so, the one great thing about this is you get information in the click of a finger. I don't see anything wrong with constantly sourcing information. And then in terms of something as important as your baby, like this is my first and only kid. I want to know, you know, and I want to sort of make the best choices and decisions that I possibly can so you aim at a healthy direction.*

Mothers particularly appreciated that they could use the internet as a way to quickly *crowdsource* or gather multiple viewpoints when trying to make a decision regarding their infant. One mother compared the unlimited amount of information on the internet with her mother, who was a single source of information:

*I really like gathering information, and the Internet's kind of an unlimited way as a source of information. And you can go to lots of sources at one time, versus like, my mom is one and it's limited.*

The internet was also considered a way to obtain information anonymously, either because mothers did not want to ask questions that might have an obvious answer or because the questions were too personal to ask someone:

*That blog has a lot of questions and answers. You can post things kind of anonymously so you don't kind of feel like I'm asking a stupid question. Like you can put like, "Is it normal for babies to have a green poop?"*

*Also, there are some things I don't want to ask friends that might feel more personal.*

**Table 2.** Themes and subthemes about maternal use of the internet and social media.

| Themes | Subthemes |
|---|---|
| Reasons that mothers turn to the internet for parenting and health information for their infant | Unlimited information available; Anonymity; Convenience; Immediate answers; Faster than information from a health care professional; Trustworthiness of the information; Up to date; Use to confirm information obtained from other sources |
| Cautionary advice about the internet | Wealth of information is overwhelming; Many nonreputable websites; Generalizability of the information |
| Reasons that mothers turn to social media for parenting and health information | Immediate affirmation and support; Honest answers; Acts as a support group; Tailored information; Trustworthiness of information |

As mothers often did not want to bother a health care professional or another person with all of their questions, they found the internet to be a convenient place to obtain information, particularly information that was not urgent. They could often find that information much more quickly than they would have if they had contacted a health care provider:

*And then some things, there might be medical things. I must have called my doctor's office like twice a week, every week, for like nine months. They go, "Hi again." At some point I'm like, I'm wearing out my poor nurses.*

*Because sometimes [going to the internet] is easy. If you're just sitting there and it's a question that's not...life or death...like, what colors should babies' poop be? Like, you're going to get the right answer and the Internet's going to get that right. I don't think that somebody's going to be wrong about it. So, for stuff like that I feel like there's no reason to bother somebody else with that kind of question.*

Mothers described the information on websites as generally trustworthy. They trusted that it was more current than the information that they received from trusted family members and friends:

*I read some of them on the Internet if there are questions about behavioral changes or developmental stages, because I think a lot of older resources like my parents, maybe they've forgotten what that stuff is like.*

Thus, they tended to use the internet to confirm the information obtained from other sources and for reassurance that their child was normal:

*[I go to the internet] just for confirmation. Just to see... There may be something else. There may be something else that my mom doesn't know.*

*Yeah. So, [the internet] just gives me a heads up on what things to look out for and if there's any concerns, like if my baby isn't smiling on a social level or hasn't at least rolled over by now, so I know, okay, I need to consult with my social network regarding what I'm seeing.*

## Cautionary Advice About the Internet

When mothers described their information sources on the internet, they cautioned each other to be careful about using only reputable websites for information. Websites such as Baby Center and WebMD, and search engines such as Google, are among those cited by mothers as being reputable sources:

*If we have a question about something, [then I can] Google, "why does my baby whatever?" And then I'll sort of look at what comes up, and I'll go to what I feel is the most reputable source on there and sort of read what they have to say about it...There's usually somebody out there who has something to say about the question. It's amazing what you can Google and somebody's asked it before.*

*I will say, do your research and just find...for instance, Google, WebMD, Baby Center, kind of read*

> *what people are saying, what the doctors are saying online. And then find a reliable source through the Internet if you don't have someone physically there. If it gave me good advice before, I'll go back to it again.*

In addition, mothers also described that the wealth of information could also become overwhelming, particularly when there were multiple differing viewpoints:

> *Everyone has a million and one opinions and every[one] has a different expert advice.*

Some mothers found the information on the internet to be so general as to be unhelpful to their specific situation. Mothers wanted the websites to be tailored to their situation and, occasionally, to their own race or ethnicity, to be assured about the websites' reliability:

> *I think especially with having a little black baby...The medical field...it's geared toward other folks and their kids...If he has cradle cap, the pictures I'm seeing are not little babies that look like my baby.*

### Maternal Perceptions of Social Media

With regard to social media, such as social network sites, email listservs, blogs, and mobile phone apps, mothers frequently used these as sources of parenting and health information and were generally extremely positive about the information that they found on these sites. They appreciated the immediacy of the affirmation and support that they found in these forums. Mothers described situations in which they asked a question about negative child-rearing experiences and received what they perceived to be honest answers:

> *People are open about [difficult experiences], but you have to ask them about it. Nobody wants to talk about like, "Oh, you have a beautiful baby. Wait until this happens, because it's terrible." And so, I don't think people talk about it. But once you bring it up, people are like, "Oh, yeah. We dealt with terrible diaper rash or we dealt with bad breastfeeding," or whatever it is...once you start talking to people, they're very willing to talk about it.*

In this way, these forums often served as virtual support groups for the mothers, providing affirmation about their situation and the infant:

> *You like to think your baby's unique. [But] every baby has the same issues when it comes down to it and so somebody's had to deal with it at some point.*

Many mothers used social media forums, specifically email listservs, as a way to meet other mothers and share experiences:

> *And over the years I've been using those more and more, joining more and more listservs. And then I came to find out that there were some moms' groups, specifically in [your town]...and they have cohorts based on the week your child was born, and you meet for several weeks, and then those women become friends.*

Mothers who had previous experience with child-rearing remarked on how the explosion of social media websites has changed how they ask for and receive information. These are become increasingly important as sources of information, with the information generally immediate and tailored to their situation:

> *People use Facebook groups a lot more than they did seven years ago. So, for example, just yesterday a woman posted on Facebook on our moms' group, "I just had my baby four hours ago and he's not nursing and I had a Caesarean. I know that this isn't an emergency, but I wanted the moms to weigh in on what I should be doing..." And probably about 12 of us responded right away. And I...just gave her what had happened to me...She got an immediate moms' [support group]; I wish I had that when I...was in the same situation...If [I] had known that there was this option of insta-moms on Facebook, that would've been a dream for me.*

They appreciate the discussions that occur in these forums and find them relevant and trustworthy. Some mothers prefer the advice coming from social media to their health care provider's advice:

> *I find that the listserv information is more reliable...we had these four-week-old babies...And a mom said that her baby was constipated and that the doctor's advice was to give him...a laxative, but I was like, "Oh my God." I mean, I think I felt like my heart stopped...I'm still traumatized by some poor mother following advice from a crazy pediatrician to think that you would give a baby some; it made no sense to me. So, I find if she had asked that question at the listserv, she would get better advice than following the advice of this doctor.*

Mothers who used mobile phone apps in particular liked how the information on the apps was timely and tailored to their individual situation:

> *You can tailor [the apps and websites] to your child's age and development, and that's why I appreciate that. Because sometimes the other stuff, the other sources of advice, they're just this nebulous baby advice but [not] based on what [my baby is doing].*

## *Discussion*

### Principal Findings

The internet and social media have rapidly become important influences, impacting numerous decisions on a daily basis. Opinions expressed online by strangers are as, or potentially more, important than those of family and friends [23-26]. In a 2018 Pew survey, 11% of US adults stated that they changed their opinion because of what they saw on social media in the past year [27]. Although these statistics are not specific to health-related decision making, our findings suggest that maternal decisions related to infant care may also be strongly influenced by what mothers access on the internet and social media.

Social learning theories [28,29] suggest that behaviors and rationalizations for those behaviors are in part learned from and reinforced by others, so that one's behavior becomes increasingly similar to that of the people to whom she/he has regular exposure. The theory of reasoned action similarly argues that volitional behavior results, in part, from subjective norms regarding a behavior, that is, one's perception of the prevailing attitudes and beliefs regarding a behavior that are held by people whom one trusts [30-32]. Our findings suggest that the internet and social media have become as influential as family and friends in modeling behavior, establishing norms, and thus shaping parenting and health-related decision making.

Mothers appreciated the fact that the internet and social media allowed them to access unlimited information instantaneously on a 24/7 basis. They did not need to wait for the next well child visit or to call their health care professional, family member, or friend for advice. They also liked the anonymity of the internet [33], which allowed them to ask questions that they might be embarrassed to ask a person.

Mothers liked to crowdsource information, that is, to gather opinions from multiple sources [9,15]. We and others have found that mothers make decisions by instinct [34,35] and/or consensus [1,15]. If there was a general agreement about the decision, then that allowed them to move forward with that decision with more confidence. Results from this study suggest that the internet, with its access to vast numbers of opinions, provided this consensus and confidence [35].

Mothers believed that information on the internet is generally trustworthy, with the caveat that you have to use reliable websites and apps. Although we did not ask them to define how they determine whether a website is reliable, they commented that previous good experience with a source increased their trust in it and that they generally went back repeatedly to those websites/apps with which they had positive previous experiences. Many volunteered specific sources that they considered trustworthy, for example, WebMD and Baby Center. Although it is not technically a website, Google was a frequently cited search engine. However, parenting information found in websites that rank highly in Google search results may often be unsupported by evidence [36]. One study found that, when it came to information about infant sleep safety, government websites were the most accurate and blogs the least [36]. Therefore, the fact that no women in our sample explicitly cited concerns about information on blogs or other crowdsourcing websites, despite the fact that these sites are frequently consulted, is concerning.

Indeed, although mothers in our study cautioned that only reliable websites should be trusted, this concern was not expressed about social media. Perhaps because these Facebook groups, blogs, and listservs were geared toward parents and because entries and responses were written by those perceived by mothers to be *just like me*, mothers in our study considered the information and opinions expressed on social media as being trustworthy, perhaps even more trustworthy than those of health care professionals. Bernhardt, in a qualitative study of southwestern US parents, found that mothers highly trusted internet content written by other parents but only in specific questions [37]. Kallem found that, in a Facebook group of mothers, peer responses to mothers' direct questions about their infants' health generally did not contradict American Academy of Pediatrics (AAP) recommendations. However, mothers' posts describing their infants' sleep practices and screen time practices frequently were not consistent with AAP recommendations [38]. It is thus concerning that this information is so highly trusted. Although there is no published research specifically on the effectiveness of health care providers guiding parents to access vetted websites and more monitored social media sites, health care providers are trusted sources of health information for many parents [39,40]. Thus, parents may appreciate their guidance regarding appropriate internet sources of parenting and health information, and this guidance may in turn improve the safety of infant care practices.

## Limitations

We acknowledge that this study has several limitations. First, our study population was limited to a single geographic region, to mothers, and to those who self-identified as AA or Caucasian. Second, although qualitative research can provide an insight into a broad array of opinions, it cannot be used to determine the prevalence of any one viewpoint. Thus, although we reached thematic saturation, our results are not necessarily generalizable to fathers, other groups, or geographic regions. However, our findings are consistent with other qualitative studies [34,35,37]. Nonetheless, further study in other geographic and racial/ethnic groups will be important to determine if these perceptions and opinions are widespread.

## Conclusions

The internet and social media are becoming important sources of health information that mothers turn to when making infant care decisions, and for some mothers, these electronic resources are more trusted than family members, friends, and health care professionals. It is becoming increasingly important that parents be provided guidance about accessing trustworthy, evidence-based health information. In addition, health care providers will need to be proactive in harnessing social media to encourage healthy decisions.

## Acknowledgments

This project was supported by the National Institute for Minority Health and Health Disparities 1R01MD007702. The study sponsor had no role in study design; collection, analysis, and interpretation of data; writing the report; and the decision to submit the report for publication. The authors have no other funding sources.

## Authors' Contributions

RM made substantial contributions to the conception of the work, analysis, and interpretation of the data and drafted the manuscript. RC made substantial contributions to the conception of the work, analysis, and interpretation of the data and revised the manuscript critically for important intellectual content. AM made substantial contributions to the acquisition and interpretation of data and revised the manuscript critically for important intellectual content. RO made substantial contributions to the acquisition and interpretation of data and revised the manuscript critically for important intellectual content.

All authors have approved the final version of the manuscript and agree to be accountable for all aspects of the study in ensuring that questions related to the accuracy or integrity of any part of the study are appropriately investigated and resolved.

## Conflicts of Interest

None declared.

## References

1. Oden R, Joyner B, Ajao TI, Moon R. Factors influencing African American mothers' decisions about sleep position: a qualitative study. J Natl Med Assoc 2010 Oct;102(10):870-2, 875-80. [doi: 10.1016/s0027-9684(15)30705-7] [Medline: 21053701]
2. Swanson V, Power KG. Initiation and continuation of breastfeeding: theory of planned behaviour. J Adv Nurs 2005 May;50(3):272-282. [doi: 10.1111/j.1365-2648.2005.03390.x] [Medline: 15811106]
3. Scott JA, Binns CW, Oddy WH, Graham KI. Predictors of breastfeeding duration: evidence from a cohort study. Pediatrics 2006 Apr;117(4):e646-e655. [doi: 10.1542/peds.2005-1991] [Medline: 16585281]
4. Brunson EK. The impact of social networks on parents' vaccination decisions. Pediatrics 2013 May;131(5):e1397-e1404. [doi: 10.1542/peds.2012-2452] [Medline: 23589813]
5. Fox S, Duggan M. Pew Research Center. Washington, DC; 2013 Jan 15. Health online 2013 URL: https://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-06-12]
6. Khoo K, Bolt P, Babl FE, Jury S, Goldman RD. Health information seeking by parents in the Internet age. J Paediatr Child Health 2008;44(7-8):419-423. [doi: 10.1111/j.1440-1754.2008.01322.x] [Medline: 18564080]
7. van der Gugten AC, de Leeuw RJ, Verheij TJ, van der Ent CK, Kars MC. E-health and health care behaviour of parents of young children: a qualitative study. Scand J Prim Health Care 2016 Jun;34(2):135-142 [FREE Full text] [doi: 10.3109/02813432.2016.1160627] [Medline: 27063729]
8. Merriam-Webster. URL: https://www.merriam-webster.com/ [accessed 2019-06-12]
9. Duggan M, Lenhart A, Lampe C, Ellison N. Pew Research Center. Washington, DC; 2015. Parents and social media URL: https://www.pewinternet.org/2015/07/16/parents-and-social-media/ [accessed 2019-06-12]
10. Baker B, Yang I. Social media as social support in pregnancy and the postpartum. Sex Reprod Healthc 2018 Oct;17:31-34. [doi: 10.1016/j.srhc.2018.05.003] [Medline: 30193717]
11. Antonucci TC, Akiyama H. An examination of sex differences in social support among older men and women. Sex Roles 1987 Dec;17(11-12):737-749 [FREE Full text] [doi: 10.1007/BF00287685]
12. Kendler K, Myers J, Prescott C. Sex differences in the relationship between social support and risk for major depression: a longitudinal study of opposite-sex twin pairs. Am J Psychiatry 2005 Feb;162(2):250-256. [doi: 10.1176/appi.ajp.162.2.250] [Medline: 15677587]
13. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. Br Med J 2000 Jan 01;320(7226):50-52 [FREE Full text] [doi: 10.1136/bmj.320.7226.50] [Medline: 10617534]
14. Krueger R, Casey M. Focus Groups: A Practical Guide For Applied Research. Thousand Oaks, CA: Sage Publications, Inc; 2019.
15. Moon RY, Mathews A, Oden R, Carlin R. A qualitative analysis of how mothers' social networks are established and used to make infant care decisions. Clin Pediatr (Phila) 2019 Apr 25:9922819845332 Epub ahead of print. [doi: 10.1177/0009922819845332] [Medline: 31018675]
16. Ajao TI, Oden RP, Joyner BL, Moon RY. Decisions of black parents about infant bedding and sleep surfaces: a qualitative study. Pediatrics 2011 Sep;128(3):494-502 [FREE Full text] [doi: 10.1542/peds.2011-0072] [Medline: 21859921]
17. Joyner BL, Oden R, Ajao TI, Moon R. Where should my baby sleep: a qualitative study of African American infant sleep location decisions. J Natl Med Assoc 2010 Oct;102(10):881-889. [doi: 10.1016/s0027-9684(15)30706-9] [Medline: 21053702]
18. Moon RY, Oden RP, Joyner BL, Ajao TI. Qualitative analysis of beliefs and perceptions about sudden infant death syndrome in African-American mothers: implications for safe sleep recommendations. J Pediatr 2010 Jul;157(1):92-97.e2. [doi: 10.1016/j.jpeds.2010.01.027] [Medline: 20303505]
19. QSR International. Melbourne, Australia: QSR International Pty Ltd; 2015. NVivo 11 URL: http://www.qsrinternational.com/nvivo/support-overview/downloads/nvivo-11-for-windows [accessed 2019-06-12]
20. Richards L. Readme First For A User's Guide To Qualitative Methods. Thousand Oaks, CA: Sage Publications, Inc; 2019.

21. Denzin N, Lincoln Y, editors. Strategies of Qualitative Inquiry. 4th edition. Thousand Oaks, CA: Sage Publications, Ltd; 2012.
22. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, And Mixed Methods Approaches. Thousand Oaks, CA: Sage Publications, Inc; 2019.
23. Wang J, Chang C. How online social ties and product-related risks influence purchase intentions: a Facebook experiment. Electron Commer Res Appl 2013 Sep;12(5):337-346. [doi: 10.1016/j.elerap.2013.03.003]
24. Bond RM, Fariss CJ, Jones JJ, Kramer AI, Marlow C, Settle JE, et al. A 61-million-person experiment in social influence and political mobilization. Nature 2012 Sep 13;489(7415):295-298 [FREE Full text] [doi: 10.1038/nature11421] [Medline: 22972300]
25. Ahmed M. Social Media Today. Is social media the biggest influencer of buying decisions? URL: https://www.socialmediatoday.com/marketing/masroor/2015-05-28/social-media-biggest-influencer-buying-decisions [accessed 2019-06-12]
26. Ye QA, Law R, Gu B, Chen W. The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Comput Hum Behav 2011 Mar;27(2):634-639. [doi: 10.1016/j.chb.2010.04.014]
27. Bialik K. Pew Research Center. Washington DC; 2018 Aug 15. 14% of Americans have changed their mind about an issue because of something they saw on social media URL: https://www.pewresearch.org/fact-tank/2018/08/15/14-of-americans-have-changed-their-mind-about-an-issue-because-of-something-they-saw-on-social-media/ [accessed 2019-06-12]
28. Akers RL, Krohn MD, Lanza-Kaduce L, Radosevich M. Social learning and deviant behavior: a specific test of a general theory. Am Sociol Rev 1979 Aug;44(4):636-655. [Medline: 389120]
29. Bandura A. Social Learning Theory. Englewood, NJ: Prentice Hall; 1977.
30. Ajzen I, Fishbein M. Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.
31. Fishbein M, Ajzen I. Belief, attitude, intention, and behavior: an introduction to theory and research. Reading, MA: Addison-Wesley; 1975.
32. Fishbein M. A reasoned action approach to health promotion. Med Decis Making 2008;28(6):834-844 [FREE Full text] [doi: 10.1177/0272989X08326092] [Medline: 19015289]
33. Cline RJ, Haynes KM. Consumer health information seeking on the internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
34. Lewallen LP. Healthy behaviors and sources of health information among low-income pregnant women. Public Health Nurs 2004;21(3):200-206. [doi: 10.1111/j.0737-1209.2004.021302.x] [Medline: 15144364]
35. Criss S, Woo BJ, Goldman RE, Perkins M, Cunningham C, Taveras EM. The role of health information sources in decision-making among Hispanic mothers during their children's first 1000 days of life. Matern Child Health J 2015 Nov;19(11):2536-2543 [FREE Full text] [doi: 10.1007/s10995-015-1774-2] [Medline: 26122256]
36. Chung M, Oden RP, Joyner BL, Sims A, Moon RY. Safe infant sleep recommendations on the internet: let's Google it. J Pediatr 2012 Dec;161(6):1080-1084 [FREE Full text] [doi: 10.1016/j.jpeds.2012.06.004] [Medline: 22863258]
37. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 1;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
38. Kallem S, Gruver RS, Virudachalam S, Fiks AG. Mothers' Facebook posts about infant health: findings from the Grow2Gether study. BMC Pediatr 2018 Dec 31;18(1):341 [FREE Full text] [doi: 10.1186/s12887-018-1315-4] [Medline: 30382827]
39. Freed GL, Clark SJ, Butchart AT, Singer DC, Davis MM. Sources and perceived credibility of vaccine-safety information for parents. Pediatrics 2011 May;127(Suppl 1):S107-S112. [doi: 10.1542/peds.2010-1722P] [Medline: 21502236]
40. Hwang SS, Rybin DV, Heeren TC, Colson ER, Corwin MJ. Trust in sources of advice about infant care practices: the SAFE study. Matern Child Health J 2016 Dec;20(9):1956-1964. [doi: 10.1007/s10995-016-2011-3] [Medline: 27129949]


## Abbreviations

**AA:** African American
**AAP:** American Academy of Pediatrics

*Edited by G Eysenbach; submitted 05.04.19; peer-reviewed by S Kallem, A Barrett, M Yap; comments to author 31.05.19; revised version received 03.06.19; accepted 03.06.19; published 09.07.19.*

*Please cite as:*
*Moon RY, Mathews A, Oden R, Carlin R*
*Mothers' Perceptions of the Internet and Social Media as Sources of Parenting and Health Information: Qualitative Study*
*J Med Internet Res 2019;21(7):e14289*
*URL: https://www.jmir.org/2019/7/e14289/*
*doi:10.2196/14289*
*PMID:31290403*

Original Paper

# The Search for Consumers of Web-Based Raw DNA Interpretation Services: Using Social Media to Target Hard-to-Reach Populations

Tiernan J Cahill[1], MA; Blake Wertz[1], MA; Qiankun Zhong[1], MA; Andrew Parlato[1], MA; John Donegan[1], MA, MBA; Rebecca Forman[1], MA; Supriya Manot[1], MA; Tianyi Wu[1], MA; Yazhu Xu[1], MA; James J Cummings[1], PhD; Tricia Norkunas Cunningham[2], MPH, MA; Catharine Wang[2], PhD, MSc

[1]Division of Emerging Media, Boston University College of Communication, Boston, MA, United States

[2]Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, United States

**Corresponding Author:**
Catharine Wang, PhD, MSc
Department of Community Health Sciences
Boston University School of Public Health
801 Massachusetts Avenue
Boston, MA, 02118
United States
Phone: 1 617 358 1475
Email: clwang@bu.edu

**Related Article:**

This is a corrected version. See correction statement: http://www.jmir.org/2019/8/e15735/

## Abstract

**Background:** In recent years, there has been a proliferation of third-party Web-based services available to consumers to interpret raw DNA from direct-to-consumer genetic testing companies. Little is known about who uses these services and the downstream health implications. Identifying this hard-to-reach population of consumers for research raised questions about the most effective recruitment methods to undertake. Past studies have found that Web-based social media survey distribution can be cost-effective for targeting hard-to-reach populations, yet comparative efficacy information across platforms is limited.

**Objective:** The aim of this study was to identify the most effective Web-based strategies to identify and recruit the target population of direct-to-consumer genetic testing users who also made use of third-party interpretation services to analyze their raw genetic data. Web-based survey recruitment methods varying by social media platform and advertising method were compared in terms of cost-effectiveness and demographics of survey respondents.

**Methods:** A total of 5 Web-based survey distribution conditions were examined: 4 paid advertising services and 1 unpaid service. For the paid services, a 2x2 quasi-experimental design compared social media platforms (Facebook vs Twitter) and advertising tracking metrics (by click vs by conversion). The fifth unpaid comparison method consisted of study postings on the social media platform, Reddit, without any paid advertising. Links to identical Web-based versions of the study questionnaire were posted for 10 to 14 days for each of the distribution conditions, which allowed tracking the number of respondents that entered and completed the questionnaire by distribution condition.

**Results:** In total, 438 individuals were recruited to the study through all conditions. A nearly equivalent number of participants were recruited from paid campaigns on Facebook (n=159) and Twitter (n=167), with a smaller sample recruited on Reddit (n=112). Significantly more participants were recruited through conversion-tracking (n=222) than through click-tracking campaigns (n=104; $Z$=6.5, $P$<.001). Response rates were found to be partially driven by organic sharing of recruitment materials among social media users. Conversion tracking was more cost-effective than click tracking across paid social media platforms. Significant differences in terms of gender and age distributions were noted between the platforms and between the tracking metrics.

**Conclusions:** Web-based recruitment methods were effective at recruiting participants from a hard-to-reach population in a short time frame. There were significant differences in the effectiveness of various paid advertising techniques. Recruitment through Web-based communities also appeared to perform adequately, yet it may be limited by the number of users accessible in open community groups. Future research should evaluate the impact of organic sharing of recruitment materials because this appeared to play a substantial role in the observed effectiveness of different methods.

***(J Med Internet Res 2019;21(7):e12980)*** doi:10.2196/12980

**KEYWORDS**

research subject recruitment; social media; survey methods; data collection methods; advertising as topic; algorithms

## Introduction

To date, there have been a number of inquiries into using social media for research recruitment and there has been little consensus in terms of results. A systematic review of 30 existing studies on social media recruitment found mixed evidence with regard to the efficacy of survey recruitment on social media but did find that such methods were consistently found to be effective when specifically targeting hard-to-reach populations—those that are difficult to find or involve in research and public health programs because of their geographical location or socioeconomic situation [1,2]. However, the review also suggested that this methodology has not been studied often enough to generate firm conclusions as to its efficacy, arguing that further research, particularly research examining the cost-efficacy of different recruitment techniques and demographic differences in the resulting samples, is necessary [1]. This study partially fills this gap in the research by directly comparing multiple analogous advertisement recruitment methods on Twitter and Facebook along with the unpromoted posts on Reddit to recruit survey respondents from the same hard-to-reach population.

Social media has been defined, in a public health context, as websites that allow users to create profiles and use those profiles to connect and interact with other users [1]. Although there are dozens to hundreds of different forms of social media, at present, most of the documented social media recruitment efforts for population health research have used Facebook [3-9]. Facebook is a large social media platform with approximately 1.56 billion daily active users [10]. Studies have found success in reaching target audiences by sharing posts within Facebook communities, enlisting respondents in *snowball sampling* campaigns, and purchasing paid advertising on the platform targeting specific demographics. Although there have been fewer studies focusing on the platform, Twitter—a somewhat smaller platform with 126 million daily active users [11]—has also been used for recruitment purposes both through researchers tweeting and retweeting recruitment tweets [12,13] and advertisements [14].

There has been little research to compare the effectiveness of recruitment from across different social media platforms, although some studies have sought to use multiple platforms for recruitment without making direct statistical comparisons [15,16]. One study did compare 2 social media platforms (Twitter and Facebook) as well as another method (distributing quick response codes through mail) but did not use directly analogous recruitment methods across conditions or identify platform-level differences [17]. As a result, direct comparisons of relative effectiveness between the platforms themselves remain a challenge.

Survey research employing social media for participant recruitment has also yet to consider the multiple recruitment strategies available on a given platform. Popular platforms, such as Facebook, enable both *cost per click* advertisement sales that charge advertisers each time a user clicks on an advertisement and *cost per conversion* sales: Advertisers are billed on the basis of specific, predefined actions that follow from a user clicking through the advertisement, such as purchasing a product or completing a questionnaire. These tracking metrics may yield different results when it comes to reaching target audiences as well as achieving a cost-effective survey sample.

This study sought to better understand the differences in survey participant recruitment between social media platforms, as well as within-platform differences resulting from different tracking methods. Targeted survey participants were a hard-to-reach population of users of direct-to-consumer genetic testing (DTC-GT) services (eg, AncestryDNA and 23andMe) who had subsequently used third-party interpretation tools to analyze their raw genetic data. The goal of the study was to compare the cost-effectiveness as well as the demographic characteristics of the sample across different platforms and between different advertising tracking metrics.

To enable a more rigorous comparison between different social media platforms, this study conducted advertising campaigns on both Facebook and Twitter—a platform deemed to possess sufficient similarities to Facebook in terms of advertising affordances and presentation of content so that comparisons can be made. In addition, click-based and conversion-based tracking metrics were used on each platform. To allow for further comparisons across social media platforms, an additional condition contrasted unpaid posts to community message boards on Reddit with the advertising campaigns on Facebook and Twitter.

This study addressed the following questions regarding platform differences, cost-efficacy, and paid versus unpaid uses of social media in survey recruitment:

> *Q1: Among paid social media campaigns, which social media platform is most effective at generating survey responses from the hard-to-reach population of DTC-GT users who had also used third-party interpretation tools?*
>
> *Q2: Among paid social media campaigns, which advertising tracking method is most effective at generating survey responses from the hard-to-reach population of DTC-GT users who had also used third-party interpretation tools?*
>
> *Q3: Do surveys conducted via paid social media campaigns on Facebook and Twitter generate more survey responses from the hard-to-reach population*

*of DTC-GT users who had also used third-party interpretation tools compared with surveys posted on open (unpaid) Web-based communities?*

*Q4: What demographic differences exist between survey respondents who are recruited using (1) different platforms and (2) different advertising tracking methods?*

## Methods

This study compared the effectiveness and cost of different social media recruitment methodologies that comprised both paid and unpaid advertising structures across different platforms (Facebook, Twitter, and Reddit). Despite its large user base (1 billion active users), Instagram was not included because of the lack of a well-defined community of interest, which was the basis for targeting advertisements toward the relevant population. The target population for the survey was defined as US residents who had undergone genetic testing via direct-to-consumer (DTC) companies and who subsequently used third-party Web-based DNA interpretation services.

### Paid Recruitment Methods

A 2x2 factorial design was used to test the comparative effectiveness and cost of different platforms and advertising tracking metrics for paid recruitment. Facebook and Twitter were selected as the platforms to be compared based on their large US resident user bases. Both platforms have proprietary content distribution networks that distribute paid advertising content to their users. Advertised content appeared as *promoted* status cards or tweets in the news feed of the targeted users (Figures 1 and 2), intermingled with user-generated content.

There are slight differences in the way each content distribution network allows for targeting of specific user demographics. An effort was made to mirror the approach taken to targeting users across both sites. On Facebook, the potential audience of the campaign was defined as users living in the United States with an interest in 23andMe, a major DTC-GT company. Facebook targets paid advertising campaigns by identifying *interests from information users have added to their Timeline, keywords associated with the Pages they like or apps they use, ads they have clicked on, and other similar sources* [18]. All users aged 18 years and older were included, resulting in a potential audience for the campaign of 740,000 Facebook users at the time of launch.

On Twitter, the potential audience of the campaign was defined as US-based Twitter users who were followers of @23andMe's Twitter account, as well as users with interests similar to followers of @23andMe. According to Twitter, "[Follower targeting] works by displaying your Twitter Ads campaigns to people who follow specific usernames or are similar to the followers of those usernames" [19]. This resulted in a potential audience size for the campaign of between 178,000 and 267,000 Twitter users at the time of launch.

A total of US $1000 was budgeted for the paid campaigns, divided evenly between Facebook and Twitter. Automatic bidding was the default on both platforms and was used in all 4 conditions. This feature dynamically adjusts the cost of advertising based on availability and demand, as well as the bidding parameters set by other advertisers. There are minor differences in the way each advertising platform handles bidding for advertisements: Twitter requires advertisers to specify a daily budget and provides an optional total budget setting for automatic bidding, after which the campaign will end. The daily budget for each condition was set at US $25 per day with a total budget of US $250 per campaign. Facebook does not require a daily budget setting; however, the total budget for this campaign was also set at US $250. Each campaign was allowed to run until the total budget was exhausted: Twitter advertisements ran for 10 days each, whereas Facebook advertisements were displayed for 14 days each. Although both Facebook and Twitter provide advertisers some control over the time of the day when advertisements are displayed, it was not specified on either platform in this study.

On both social media platforms, 2 advertising campaigns were conducted using different payment structures corresponding to different tracking methods. Both platforms allow the advertiser to either pay for each click through to the advertiser's landing page (cost per click) or to pay for each iteration of a defined conversion action after the user has clicked through to the landing page (cost per conversion). For the purposes of the study, a *conversion* was defined as the user reaching the end of the questionnaire.

Both advertising platforms claim to iteratively *optimize* the targeting of a given advertising campaign based on the tracking metric used. Thus, a campaign for which the advertiser is billed per click is purportedly targeted in such a way as to maximize the likelihood that a given user who is shown the content will click on it. Conversely, a campaign for which the advertiser is billed per conversion is purportedly targeted in such a way as to maximize the likelihood that a given user will complete the conversion action after having clicked through.

To track which of the users that were shown the recruitment material ended up completing the questionnaire and allowing feedback to the content distribution network for optimization purposes, a tracking pixel was used for the 2 campaigns in the conversion-based condition. A tracking pixel is a hidden image file embedded in a custom landing page, which users were automatically redirected to after having completed the questionnaire. Loading the image in a Web browser triggers a JavaScript function on the page, which logs the conversion with either Twitter or Facebook, depending on which version of the questionnaire was completed. In contrast, the 2 campaigns using the click-based condition only tracked how many users clicked the link to the questionnaire rather than any user interaction with the questionnaire. Each campaign used a separate, yet identical, Web-based questionnaire, enabling survey respondents to be categorized by the advertising campaign that recruited them.

**Figure 1.** Example of recruitment materials on Twitter.


**Figure 2.** Example of recruitment materials on Facebook.

### Unpaid Recruitment Method

A parallel recruitment campaign was conducted on Reddit, a social news and community discussion site, to assess the viability of recruitment through unpaid posts to relevant Web-based communities. Reddit was selected because of the presence of several relevant community groups (see the table provided in Multimedia Appendix 1) as well as the open structure of the site, which allows any user to post to any public group or subreddit, subject to community moderation. In total, 13 relevant subreddits were identified, although r/Health was not used because of community guidelines that prohibited the posting of content other than news. Identical posts were made on each of the remaining 12 subreddits seeking respondents for the survey (see the textbox provided in Multimedia Appendix 1).

### Statistical Analysis

#### *Data Preparation*

The dataset was screened for duplicate responses using the internet protocol (IP) address and demographic profile of respondents, where responses from the same IP address within a 24-hour period or responses from the same IP address with a matching demographic profile were flagged as duplicates. This resulted in 17 responses being removed from the subsequent analysis. No responses were found to have been duplicated more than once, suggesting that these were likely the result of user error rather than a systematic effort.

#### *Recruitment Effectiveness*

A chi-square test was conducted to determine the extent to which the proportion of observed frequencies among the 4 paid campaigns conformed to a discrete uniform distribution, which would suggest the absence of a measurable difference in recruitment effectiveness between conditions. Posthoc pairwise *Z* tests were performed between all campaigns with a Bonferroni correction for multiple comparisons.

#### *Cost-Effectiveness*

The recruitment budget for each condition was fixed at US $250, such that more cost-effective methods would yield a greater total number of responses over the study period. On the basis of this fixed budget, the cost-efficacy of each paid campaign was calculated in terms of the cost per survey response and cost per 1000 impressions. Each *impression* marks a time when the recruitment materials were displayed to a user, regardless of whether that user had seen the materials before or interacted with them in any way.

#### *Survey Demographics*

A total of 4 demographic variables were collected in the survey: age, gender, education, and race and ethnicity. Chi-square tests of homogeneity were performed to determine the statistical significance of differences in the distributions of gender and ethnicity. Participants who reported their gender as neither male nor female were excluded from the analysis of gender distributions because of the absence of reliable information on the expected proportion of nonbinary gender identifying individuals in the population. Age distributions were compared using a one-way analysis of variance. Posthoc pairwise comparisons were conducted with a Bonferroni correction, as appropriate. Kruskal-Wallis H tests were used to compare the education level of respondents. To maximize the response rate, demographic questions were not required to complete the survey. For demographic analyses only, participants who did not report the demographic characteristic of interest were excluded.

## Results

### Study Participants

Participant demographics are presented in Table 1. Notably, because demographic questions were optional in the questionnaire, a substantial portion of respondents who completed the rest of the questionnaire elected not to answer them.

The mean age of those who reported this (n=266) was 46 years at the time of the survey. Among respondents who reported their gender (n=298), the majority (204/298, 68.5%) were female. The median level among those who reported their level of education (n=296) was a 4-year college degree across all conditions. Among respondents who reported their race or ethnicity (n=294), the majority (238/294, 81.0%) were white.

**Table 1.** Participant demographics.

| Demographic variables | Participants (N=438), n (%) | Participants (excluding missing), n (%) |
|---|---|---|
| **Age (years)** | | **N=266** |
| 18-24 | 24 (5.5) | 24 (9.0) |
| 25-44 | 101 (23.1) | 101 (38.0) |
| 45-64 | 109 (24.9) | 109 (41.0) |
| 65 and older | 32 (7.3) | 32 (12.0) |
| Did not report | 172 (39.3) | —[a] |
| **Gender** | | **N=298** |
| Female | 204 (46.6) | 204 (68.5) |
| Male | 93 (21.2) | 93 (31.2) |
| Other | 1 (0.2) | 1 (0.3) |
| Did not report | 140 (32.0) | — |
| **Education** | | **N=296** |
| Less than high school | 3 (0.7) | 3 (1.0) |
| High school/GED[b] | 11 (2.5) | 11 (3.7) |
| Some college | 63 (14.4) | 63 (21.3) |
| 2-year college degree | 39 (8.9) | 39 (13.2) |
| 4-year college degree | 91 (20.8) | 91 (30.7) |
| Advanced degree (postgraduate) | 89 (20.3) | 89 (30.1) |
| Did not report | 142 (32.4) | — |
| **Race and Ethnicity** | | **N=294** |
| White/Caucasian | 238 (54.3) | 238 (81.0) |
| African American | 7 (1.6) | 7 (2.4) |
| Hispanic/Latino | 10 (2.3) | 10 (3.4) |
| Asian | 8 (1.8) | 8 (2.7) |
| Multiethnic | 23 (5.3) | 25 (7.8) |
| Other | 8 (1.8) | 8 (2.7) |
| Did not report | 144 (32.9) | — |

[a]Valid percentage excludes respondents who did not report for a given demographic variable.

[b]GED refers to those respondents who reported completing the General Education Development tests as their highest level of educational attainment.

## Recruitment Effectiveness

A total of 540 responses were received in the survey; however, 17 duplicate responses were identified during data cleaning, and an additional 88 respondents did not report having used a DTC-GT or did not report being aware of any third-party genetic interpretation companies and were subsequently excluded from the final sample (*N*=438). There were significant differences in the frequency of survey responses between the different experimental conditions ($\chi^2_3$=84.2; *P*<.001). See Figure 3 for frequencies and Table 2 for pairwise comparisons. Nearly equal samples were collected from paid campaigns on Facebook (159/438, 36.3%) and Twitter (167/438, 38.1%). A significant but somewhat smaller sample of participants was recruited through the parallel unpaid campaign on Reddit (112/438, 25.6%). Of the participants recruited through paid campaigns (n=326), significantly more were recruited through the conversion-tracking campaigns (222/326, 68.1%) than through the click-tracking campaigns (104/326, 31.9%; *Z*=6.5; *P*<.001). The difference between conversion-based and click-based tracking metrics was much more pronounced on Twitter than on Facebook (*Z*=6.7, *P*<.001); correspondingly, of the 5 recruitment methodologies used, the Twitter-Conversion campaign recruited the greatest number of participants (142/438, 32.4%) and the Twitter-Click campaign recruited the fewest (25/438, 5.7%).

**Figure 3.** Respondent count by recruitment platform and tracking metric.

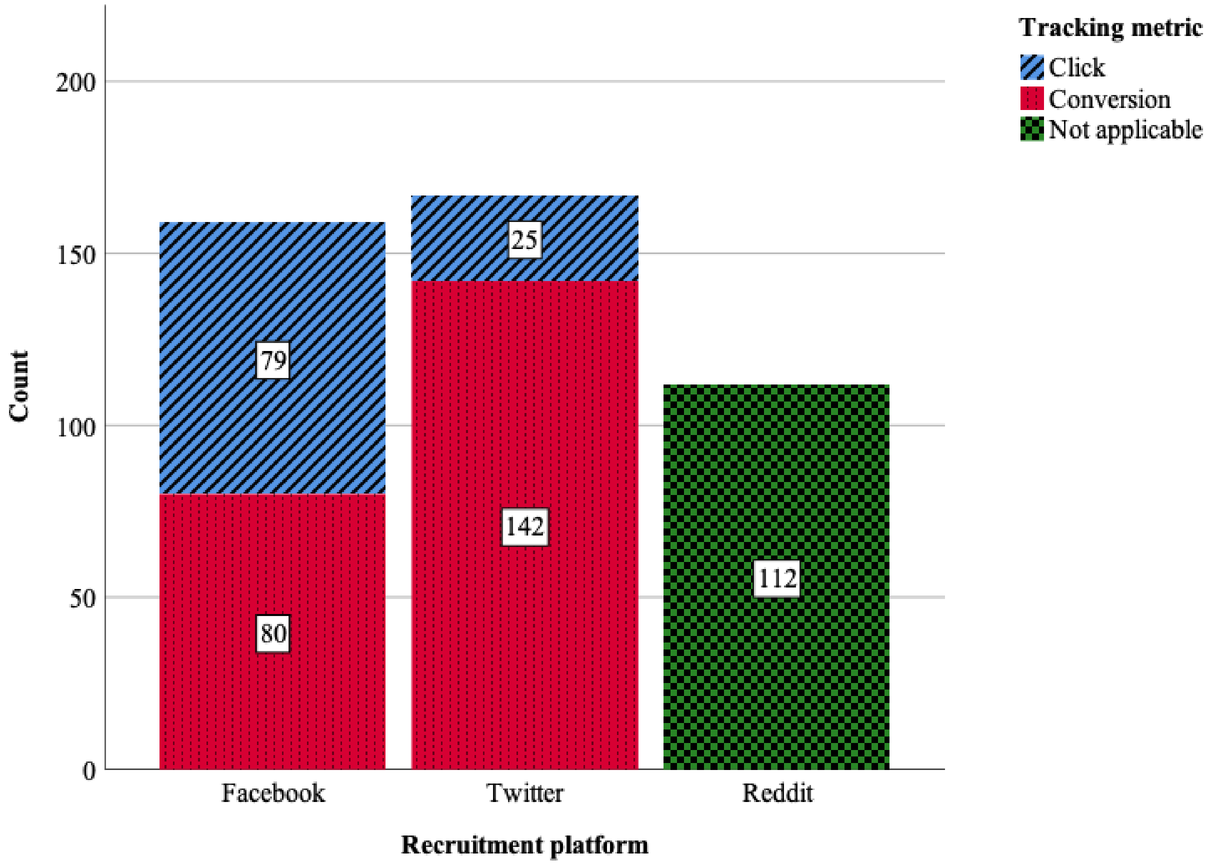



**Table 2.** Pairwise tests comparing the number of respondents recruited in paid conditions (Bonferroni-corrected threshold for statistical significance is alpha=.01).

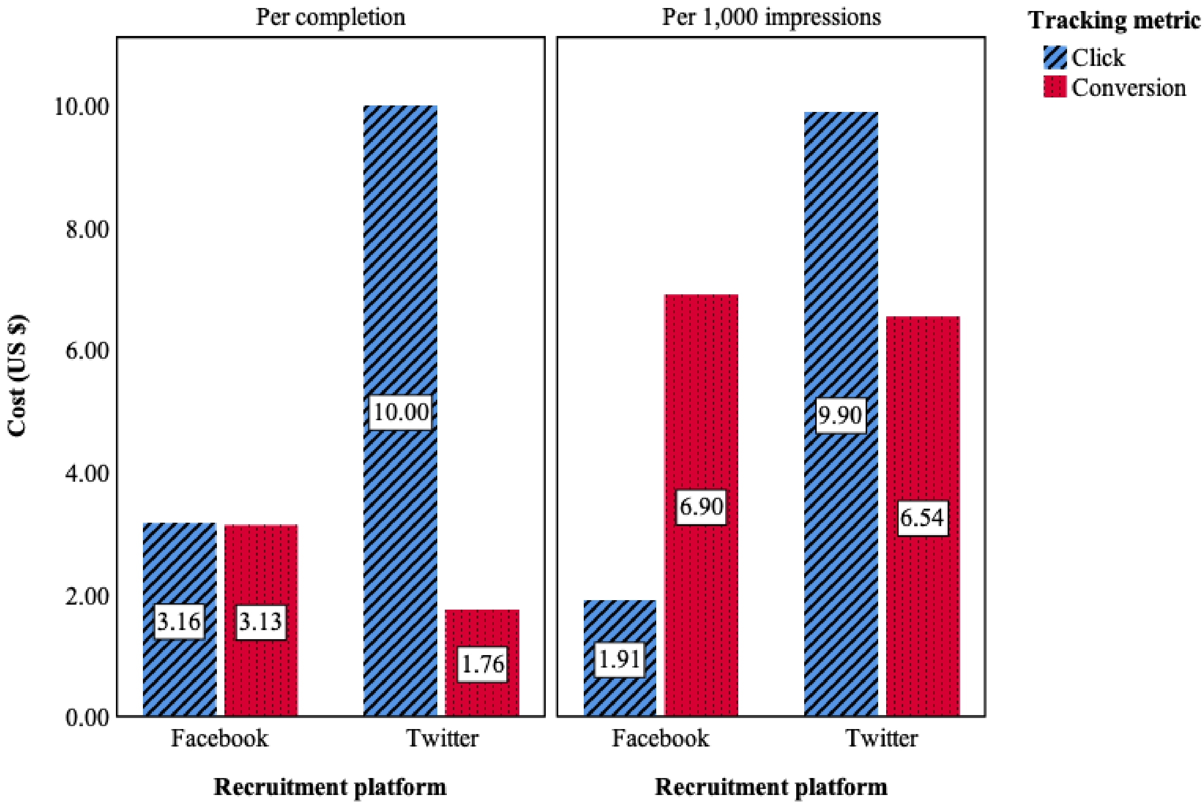
| Condition 1 (n) | Condition 2 (n) | *Z* score | *P* value |
|---|---|---|---|
| Facebook-Click (79) | Twitter-Click (25) | 5.3 | <.001 |
| Facebook-Conversion (80) | Twitter-Conversion (142) | 3.3 | <.001 |
| Facebook-Click (79) | Twitter-Conversion (142) | 4.3 | <.001 |
| Facebook-Conversion (80) | Twitter-Click (25) | 5.4 | <.001 |
| Twitter-Click (25) | Twitter-Conversion (142) | 9.1 | <.001 |
| Facebook-Click (79) | Facebook-Conversion (80) | 0.1 | .94 |

## Cost-Effectiveness

Conversion-tracking campaigns on both Facebook and Twitter were more cost-effective at garnering survey respondents, averaging US $3.13 and US $1.76 per response, respectively. Click-based campaigns cost an average of US $3.16 and US $10.00 per response on the same platforms. There was a substantial difference in the cost of impressions between platforms in the click-based conditions, with Facebook charging US $1.91 per 1,000 impressions compared with US $9.90 on Twitter. There was, however, only a nominal difference in the cost of impressions among conversion-based conditions.

Of the 4 paid advertising conditions, the Twitter-Conversion campaign was the most cost-effective in terms of generating survey responses, followed by the Facebook-Conversion, Facebook-Click, and Twitter-Click campaigns. The Facebook-Click campaign was the most cost-effective in generating broad audience exposure, followed by the Twitter-Conversion, Facebook-Conversion, and Twitter-Click campaigns (Figure 4).

**Figure 4.** Cost-effectiveness measures by recruitment method.




## Demographic Comparisons Across Platforms and Tracking Methods

### *Differences Between Platforms*

There were notable differences in the demographic characteristics of survey respondents recruited on each of the 3 platforms (Facebook, Twitter, and Reddit), particularly by age, gender, and race and ethnicity. There was no significant difference in the level of education reported by respondents recruited across different platforms ($H_{2=}$ 4.61, $P$=.10).

Among those who reported their age (n=266), there were significant differences in the average age of respondents recruited on different platforms ($F_{2,263}$=58.18; $P$<.001). The age distributions for respondents recruited on each platform are presented in Table 3. There was not a significant difference between the age of respondents recruited on Facebook (mean 49.13) and Twitter (mean 53.11); however, respondents recruited on Reddit were, on average, significantly younger than either of the other 2 groups (mean 34.23).

There were significant differences in the ratio of female to male respondents between those who were recruited on different platforms ($\chi^2_2$=53.0; $P$<.001). The gender distributions for each recruitment platform are presented in Figure 5. Female respondents made up the majority on both Facebook (65/82, 79%) and Twitter (102/122, 83.6%), but were in the minority among those recruited on Reddit (37/94, 39%).

Table 4 contains a complete reporting of respondent race and ethnicity by recruitment platform. The difference in the proportion of white to nonwhite respondents across platforms approached significance ($\chi^2_2$=5.7; $P$=.06). The proportion of white respondents was higher on Twitter (105/121, 86.8%) than on Facebook (58/79, 73%) or Reddit (75/94, 80%).

**Table 3.** Respondent age by recruitment platform and tracking metric.

| Category | Age (years), mean (SD) | *F* test (*df1*,*df2*) | *P* value |
|---|---|---|---|
| **Recruitment platform (n)** | | | |
| Facebook (69) | 49.13 (12.64) | 58.18 (2,263) | <.001 |
| Twitter (110) | 53.11 (12.95) | 58.18 (2,263) | <.001 |
| Reddit (87) | 34.23 (11.89) | 58.18 (2,263) | <.001 |
| Overall (266) | 45.90 (15.00) | —[a] | — |
| **Tracking metric (n)** | | | |
| Click (50) | 51.93 (12.67) | 0.35 (1,177) | .56 |
| Conversion (129) | 51.81 (13.08) | 0.35 (1,177) | .56 |
| Overall (179) | 51.58 (12.94) | — | — |

[a]Not applicable.

**Figure 5.** Distribution of respondent gender by recruitment platform.

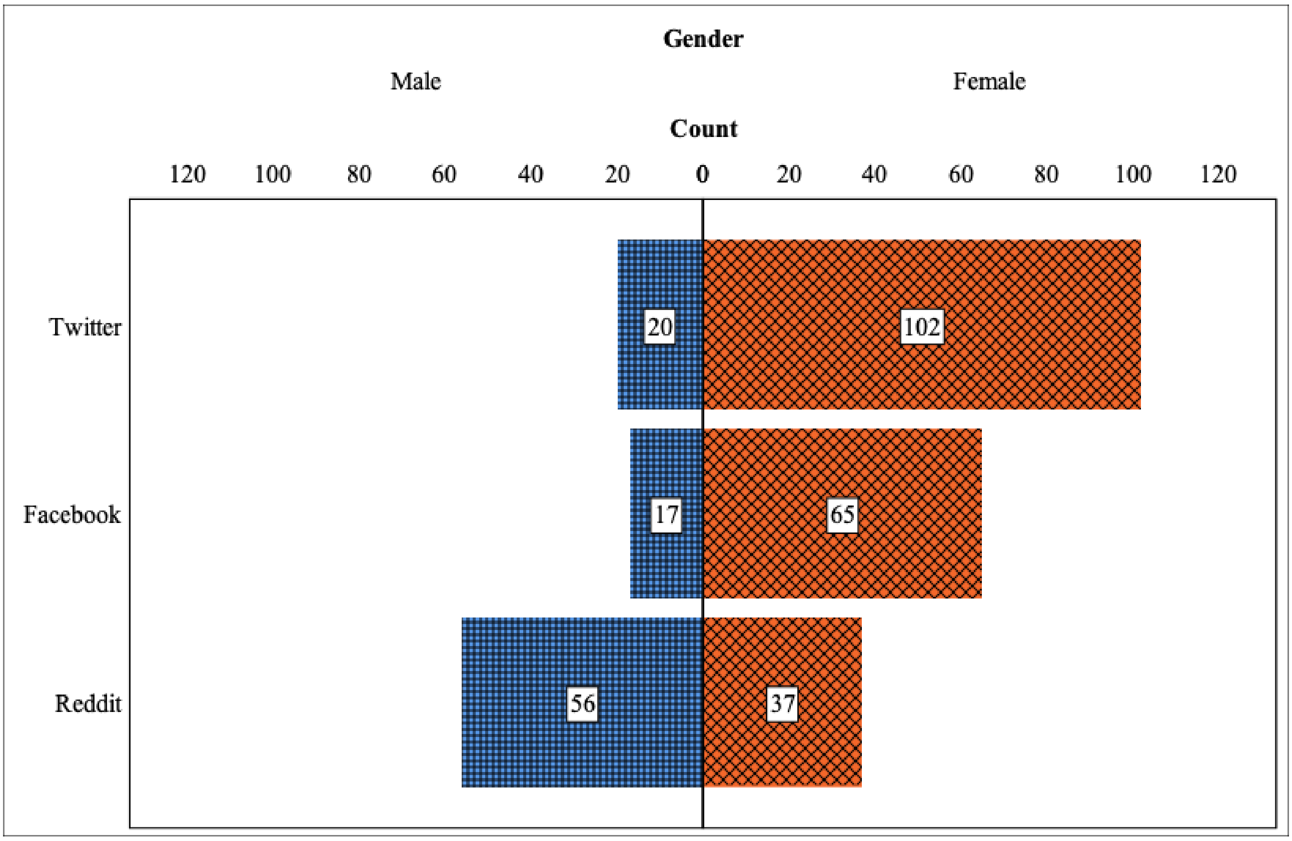
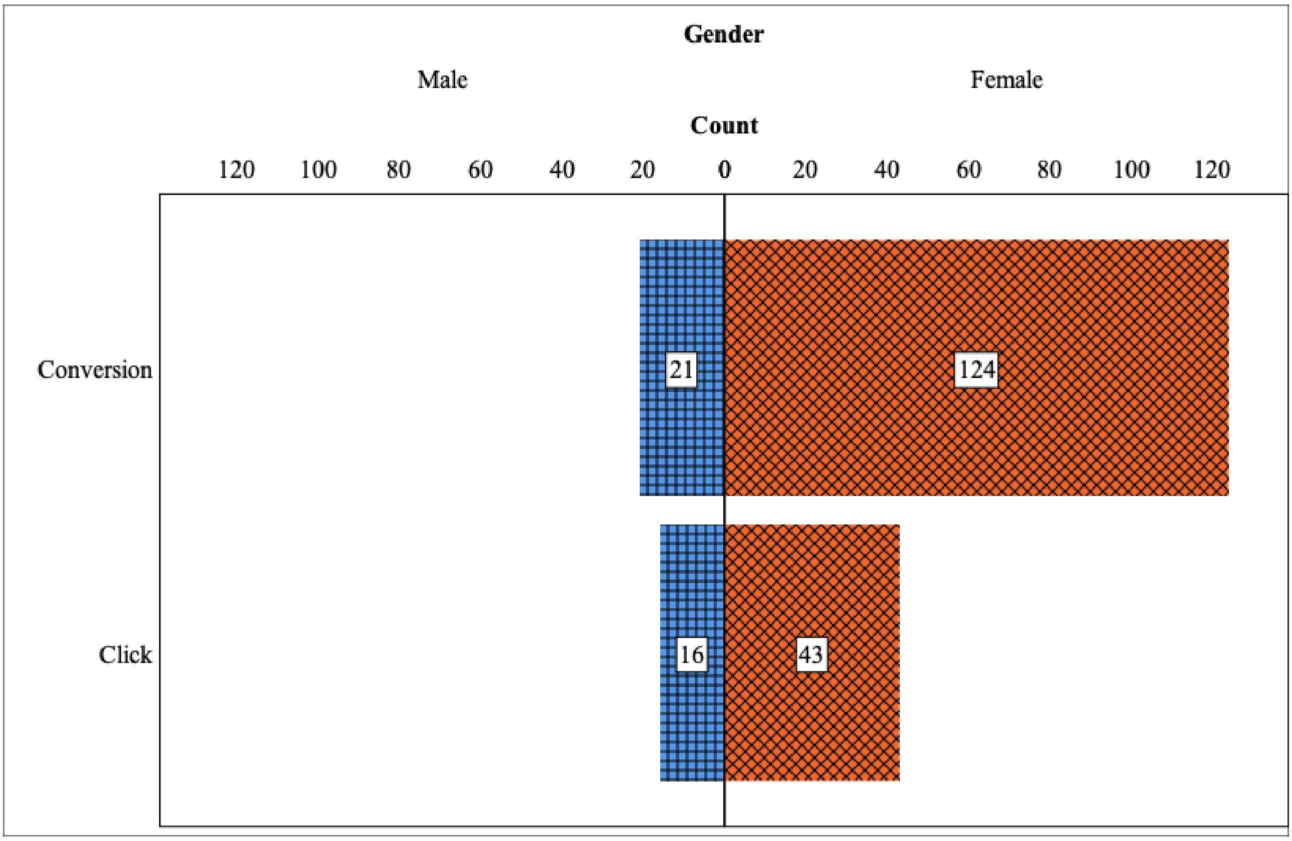


**Table 4.** Respondent race and ethnicity by recruitment platform and tracking metric.

| Category | Race and ethnicity | | | | | |
|---|---|---|---|---|---|---|
| | White/Caucasian | African American | Hispanic/Latino | Asian | Multiethnic | Other |
| **Recruitment platform** | | | | | | |
| Facebook (N=79) | 58 (73) | 2 (3) | 5 (6) | 1 (1) | 8 (10) | 5 (6) |
| Twitter (N=121) | 105 (86.8) | 3 (2.5) | 4 (3.3) | 2 (1.7) | 5 (4.1) | 2 (1.7) |
| Reddit (N=94) | 75 (80) | 2 (2) | 1 (1) | 5 (5) | 10 (11) | 1 (1) |
| Overall (N=294) | 238 (81.0) | 7 (2.4) | 10 (3.4) | 8 (2.7) | 23 (7.8) | 8 (2.7) |
| **Tracking metric** | | | | | | |
| Click (N=57) | 44 (77) | 2 (4) | 2 (4) | 0 (0) | 6 (11) | 3 (5) |
| Conversion (N=143) | 119 (83.2) | 3 (2.1) | 7 (4.9) | 3 (2.1) | 7 (4.9) | 4 (2.8) |
| Overall (N=200) | 163 (81.5) | 5 (2.5) | 9 (4.5) | 3 (1.5) | 13 (6.5) | 7 (3.5) |

[a]Respondents who selected more than 1 option for race and ethnicity.

**Figure 6.** Distribution of respondent gender by tracking metric.




## *Differences Between Tracking Methods*

The demographic characteristics of survey respondents recruited on paid advertising platforms (ie, Facebook and Twitter) using click-tracking were compared with those recruited using conversion-tracking. Respondents recruited from Reddit were excluded from this analysis, as no tracking method was used on this platform. The demographic differences between tracking methods were generally less substantial than those observed between platforms. The only significant difference noted was in the ratio of female to male respondents between those recruited using different tracking methods ($\chi^2_1$=4.5; *P*=.03). Gender distributions for each tracking method are summarized in Figure 6. Female respondents made up the majority in both cases, but were more prevalent among those recruited using conversion-tracking (124/145, 85.5%), compared with those recruited using click-tracking (43/59, 73%).

There was no significant difference in the age distribution of respondents recruited using conversion-tracking compared with click-tracking ($F_{1,177}$=0.35, *P*=.56) or in their level of education ($H_1$=0.02, *P*=.88). Similarly, no significant differences were found between tracking methods in the proportion of white to nonwhite respondents recruited ($\chi^2_1$=1.0, *P*=.32).

## Discussion

This study set out to test and evaluate the use of social media platforms as a recruitment tool for research on a hard-to-reach population. To do so, it directly compared paid and unpaid recruitment campaigns implemented on multiple social media platforms (Facebook, Twitter, and Reddit) and employed different advertising tracking metrics (click-based and conversion-based). Only a handful of studies have directly compared multiple methods of survey recruitment on social media; thus, this study represents a novel contribution to the development of Web-based survey methodology in general and recruitment approaches for hard-to-reach populations in particular.

Nearly identical sample sizes were obtained via paid Facebook (n=159) and Twitter (n=167) advertising, as well as the sample obtained via unpaid posting on Reddit (n=112). Although survey recruitment on social media for population health research has predominantly taken place on Facebook, this finding suggests that targeted advertising via other social media platforms may also be viable. Although the overall user base of Twitter is often noted to be substantially smaller than that of Facebook, and this was reflected in the reach of recruitment campaigns on that platform, this did not appear to constrain the effectiveness of Twitter as a platform for recruiting participants, with both Facebook and Twitter yielding comparable numbers of participants over both tracking metrics. Similarly, unpaid posting to community groups may also prove productive in achieving a broader sample.

The difference in the effectiveness of survey recruitment on Facebook and Twitter when considering both tracking metrics was found to be negligible; however, significant differences in the effectiveness of different tracking metrics across platforms were observed. Conversion-tracking campaigns recruited more than twice the number of respondents recruited by click-tracking campaigns, given the same budget. These results suggest that the use of different tracking metrics has important implications in determining the success of survey recruitment campaigns and warrants further investigation.

Overall, the use of conversion-tracking on Twitter was found to be the most cost-effective combination of tracking metric and platform conditions. The effectiveness of this approach may have been partially driven by organic growth (ie, individual users reposting recruitment materials from their own accounts). The number of survey responses garnered by this campaign exceeded the number of clicks detected on the advertisement, suggesting that this version of the recruitment materials was shared beyond the initial target audience for the advertising campaign. In an open-response question attached to the Reddit version of the questionnaire that asked respondents to identify the subreddit through which they had been recruited, 4 respondents indicated that recruitment materials had been forwarded to them by a friend or family member. Although these are the clearest indications for organic growth among the study conditions, it is possible that similar redistribution may have occurred in other cases as well.

This observation should serve as a reminder that all Web-based survey recruitment materials have the potential to be redistributed beyond the initial target audience, or otherwise *go viral*, unless steps are taken to prevent this. This potential may be useful in recruiting a larger sample or if recruiting entirely on social media platforms that do not allow targeted advertising. Here, researchers may wish to adopt from the existing literature on predictors of advertising message virality [20,21]. There is also significant cause for concern in contexts where nontarget audiences may be undesirable because of the scope or subject matter of the survey. Web-based surveys dealing with contentious topics have, in recent years, been redistributed in partisan discussion groups with the goal of sending a political message through the community's collective response [22,23]. Surveys regarding health issues where significant public controversy exists might likewise be subject to purposive redistribution with the intent of affecting the results, even if initially targeted at a more limited audience. Future research should seek to identify the extent of organic sharing of survey recruitment materials and distinguish between the data collected from targeted and nontargeted respondents.

Demographic differences among study participants were observed between social media platforms and tracking methods. It should be noted that observed differences in the demographic makeup of samples apply only to the subset of participants in each who answered optional demographic questions. Reddit attracted a sample that was approximately 15 years younger on average than that recruited from either Facebook or Twitter. In addition, the sample recruited from Reddit included far more male respondents than that of either Facebook or Twitter. No significant difference between platforms was observed for either education levels or race and ethnicity.

In terms of tracking methods, conversion-tracking resulted in a sample that included more women than that recruited through click-tracking. No other demographic differences were observed. It is possible, given the higher number of female respondents observed among those recruited on Facebook and Twitter, that conversion-based targeting may have skewed the sample even further toward female respondents by iteratively targeting female users at a higher rate than male users.

The demographic breakdown of survey participants closely matches that of other surveys conducted with DTC testing consumers [24,25], reflecting early adopters of this technology who are primarily white and highly educated. Although this study appeared to represent more females than males, past surveys have shown gender variation across DTC companies themselves, which may reflect demographic differences in user base [25]. Similarly, although recruitment on Reddit, compared with Facebook and Twitter, resulted in a very different set of respondents, it is unlikely that any of the samples is more intrinsically representative of anything beyond the respective platform's user base. As such, conducting recruitment on multiple platforms likely facilitated access to a more demographically diverse set of respondents, which yielded a final sample population that was more consistent with the past studies.

## Limitations

A major limitation for any study of modern Web-based advertising is the issue of algorithm dynamics or *the changes made by engineers to improve the commercial service and by consumers in using that service* [26]. Research findings on the effectiveness of specific software tools are intrinsically limited by the potential of such tools to evolve and change in unpredictable ways. The platform features used in this study to select an audience may not be viable in the near future, and Facebook or Twitter may update their advertisement targeting algorithm to select interested groups more effectively or more narrowly than researchers intend.

Limitations also arise from differences in the affordances of advertising platforms for defining a target audience as well as parameters for the display and timing of advertisements. Although the practical implications of these distinctions may be negligible, they do undermine the ability of researchers to directly compare the performance of advertising materials on different platforms or otherwise require researchers to isolate the most salient points of comparison: for example, in this study, differences in the way Facebook and Twitter handled bidding for advertising space meant that campaigns could either be restricted to equal budgets or to equivalent timeframes, with the former ultimately being deemed more relevant to the research questions at hand.

There were also substantial differences in the affordances of each platform for displaying advertising materials: the amount and size of displayed text, as well as the availability and size of graphics, are constrained both by the technological limitations of the platform (eg, Twitter's character limit) as well as community norms and expectations. This study adjusted the advertising materials displayed in each condition to best take advantage of the affordances of that platform: for example, more extensive copy was displayed to Reddit users before clicking on the recruitment link than users on Facebook or Twitter. Although this allowed the experiment to conform more closely to the norms of each platform, and thus supported its ecological validity, it does introduce further limitations on the direct comparability of results across platforms.

When conducting survey research, numerous considerations can influence the desired sample size. Although unpaid recruitment in Web-based community groups may perform comparably to recruitment via paid advertisements on social media, it should be noted that the number of potential respondents reached through paid advertising is more readily scalable, given a sufficient budget. By contrast, the potential audience in Web-based communities is limited by the number of active users, which may be quite small for hard-to-reach populations of interest. Such communities typically frown on repeated posting, further limiting the audience that may be reached to those who are active in the period immediately following the initial post. The inherent limits on the scope of populations that can be reached through Web-based communities may, therefore, render unpaid Web-based recruitment less effective than paid advertising for achieving larger sample sizes.

In addition, paid advertising platforms allowed for audiences to be targeted based on location. In cases of organic sharing of those advertisements, as well as recruitment in community groups, no similar controls are available. Future research should compare survey data from targeted audiences against those reached through organic growth. Likewise, an assessment of how particular tracking metrics may lead to targeting of particular demographic or interest groups is necessary to fully understand the implications of using Web-based advertising for survey recruitment. As audience targeting and tracking algorithms continue to develop, longitudinal sampling of the same small population may be useful to evaluate whether algorithm dynamics have significant effects on Web-based survey recruitment.

These limitations only stress the need for additional comparative studies of survey recruitment through different Web-based advertising platforms and tracking metrics. By understanding and evaluating the results of Web-based distribution, researchers can be aware of the effectiveness and limitations of various targeting and tracking approaches. Likewise, by comparing the characteristics of respondents from multiple recruitment campaigns, it is possible to test the effectiveness of the different methods in reaching target populations. The results of this study suggest that conversion-tracking metrics support more cost-effective survey recruitment than conventional designation of audience parameters accompanied by click-based tracking. However, algorithmic targeting of advertisements also poses problems for the reliability and reproducibility of survey research as sampling mechanisms may change in unpredictable ways.

## Conclusions

The results of this study indicate that there are meaningful differences between different approaches to Web-based survey recruitment. Advertisements on social media are a pragmatic method for survey recruitment, particularly within hard-to-reach populations and are most effective when combined with conversion-based tracking metrics. Recruitment through Web-based community groups is an effective complementary approach for reaching such populations and may give access to a more diverse sample than advertising alone. These tools must be used with due intentionality and an awareness of limitations so as to avoid potential pitfalls. Future research is needed to fully understand the effect of organic sharing and algorithm dynamics on the constitution of Web-based samples.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Reddit Campaign Details.

[DOCX File, 14KB - jmir_v21i7e12980_app1.docx ]

## References

1. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Dec 7;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
2. Atkinson R, Flint J. Social Research Update - University of Surrey. 2001. Accessing Hidden and Hard-To-Reach Populations: Snowball Research Strategies URL: http://sru.soc.surrey.ac.uk/SRU33.PDF [accessed 2019-06-25]
3. Thornton LK, Harris K, Baker AL, Johnson M, Kay-Lambkin FJ. Recruiting for addiction research via Facebook. Drug Alcohol Rev 2016 Dec;35(4):494-502. [doi: 10.1111/dar.12305] [Medline: 26179224]
4. King DB, O'Rourke N, DeLongis A. Social media recruitment and online data collection: a beginner's guide and best practices for accessing low-prevalence and hard-to-reach populations. Can Psychol 2014;55(4):240-249. [doi: 10.1037/a0038087]
5. Lohse B, Wamboldt P. Purposive facebook recruitment endows cost-effective nutrition education program evaluation. JMIR Res Protoc 2013 Aug 15;2(2):e27 [FREE Full text] [doi: 10.2196/resprot.2713] [Medline: 23948573]
6. Baltar F, Brunet I. Social research 2.0: virtual snowball sampling method using Facebook. Internet Res 2012 Jan 27;22(1):57-74. [doi: 10.1108/10662241211199960]
7. Baltar F, Gorjup MT. Online mixted sampling: an application in hidden populations. Intang Cap 2012 Apr 28;8(1):123-149. [doi: 10.3926/ic.294]
8. Bhutta CB. Not by the book: Facebook as a sampling frame. Sociol Methods Res 2012 Mar 21;41(1):57-88. [doi: 10.1177/0049124112440795]
9. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012 Feb 23;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
10. Investor Relations. 2019. Facebook Reports First Quarter 2019 Results URL: https://investor.fb.com/investor-news/press-release-details/2019/Facebook-Reports-First-Quarter-2019-Results/default.aspx [accessed 2019-05-23]
11. Shareholder Letter: Twitter. 2018. Q4 and Fiscal Year 2018 Letter to Shareholders URL: https://s22.q4cdn.com/826641620/files/doc_financials/2018/q4/Q4-2018-Shareholder-Letter.pdf [accessed 2019-05-23]
12. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. J Adv Nurs 2014 Mar;70(3):599-609. [doi: 10.1111/jan.12222] [Medline: 23909740]
13. Horrell LN, Kneipp SM, Gonzales CM, Williams KS. Using Social Media to Enhance Study Recruitment: How Facebook and Twitter May Engage Vulnerable Populations. In: Proceedings of the 27th International Nursing Research Congress. 2016 Presented at: NRC'16; July 21-25, 2016; Cape Town, South Africa URL: https://sigma.nursingrepository.org/handle/10755/620661
14. Guillory J, Kim A, Murphy J, Bradfield B, Nonnemaker J, Hsieh Y. Comparing Twitter and online panels for survey recruitment of e-cigarette users and smokers. J Med Internet Res 2016 Dec 15;18(11):e288 [FREE Full text] [doi: 10.2196/jmir.6326] [Medline: 27847353]
15. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014 May 1;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
16. Quach S, Pereira JA, Russell ML, Wormsbecker AE, Ramsay H, Crowe L, et al. The good, bad, and ugly of online recruitment of parents for health-related focus groups: lessons learned. J Med Internet Res 2013 Nov 14;15(11):e250 [FREE Full text] [doi: 10.2196/jmir.2829] [Medline: 24231040]
17. Gu LL, Skierkowski D, Florin P, Friend K, Ye Y. Facebook, Twitter, & QR codes: an exploratory trial examining the feasibility of social media mechanisms for sample recruitment. Comput Human Behav 2016 Jul;60:86-96. [doi: 10.1016/j.chb.2016.02.006]
18. Facebook Business. URL: https://www.facebook.com/business/products/ads/ad-targeting [accessed 2017-02-23]
19. Twitter Business. Follower Targeting on Twitter URL: https://business.twitter.com/en/targeting/follower.html [accessed 2017-02-23]
20. Alhabash S, McAlister AR. Redefining virality in less broad strokes: predicting viral behavioral intentions from motivations and uses of Facebook and Twitter. New Media Soc 2014 Feb 24;17(8):1317-1339. [doi: 10.1177/1461444814523726]
21. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. Am Psychol 2015 Sep;70(6):543-556. [doi: 10.1037/a0039210] [Medline: 26348336]
22. Reddit. 2017. The Trump Administration is Sending Out a Survey (Primarily to His Supporters) About Accountability of the Mainstream Media. Fill It Out Here! URL: https://www.reddit.com/r/BlueMidterm2018/comments/5uj54n/the_trump_administration_is_sending_out_a_survey/ [accessed 2017-07-10]

23. Reddit. 2017. Mainstream Media Accountability Survey Put Out by Trump. r/all is Trying to Swarm the Survey URL: https://www.reddit.com/r/The_Donald/comments/5ul4m1/mainstream_media_accountability_survey_put_out_by/ [accessed 2017-07-10]
24. Kaufman DJ, Bollinger JM, Dvoskin RL, Scott JA. Risky business: risk perception and the use of medical services among customers of DTC personal genetic testing. J Genet Couns 2012 Jun;21(3):413-422. [doi: 10.1007/s10897-012-9483-0] [Medline: 22278220]
25. Roberts JS, Gornick MC, Carere DA, Uhlmann WR, Ruffin MT, Green RC. Direct-to-consumer genetic testing: user motivations, decision making, and perceived utility of results. Public Health Genomics 2017;20(1):36-45 [FREE Full text] [doi: 10.1159/000455006] [Medline: 28068660]
26. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]


## Abbreviations

**DTC:** direct-to-consumer
**DTC-GT:** direct-to-consumer genetic testing
**IP:** internet protocol


*Edited by G Eysenbach; submitted 28.11.18; peer-reviewed by A Kim, K Himes, B Bie, F Timimi, J Cassaniti; comments to author 31.03.19; revised version received 24.05.19; accepted 04.06.19; published 30.07.19.*

*Please cite as:*
*Cahill TJ, Wertz B, Zhong Q, Parlato A, Donegan J, Forman R, Manot S, Wu T, Xu Y, Cummings JJ, Norkunas Cunningham T, Wang C*
*The Search for Consumers of Web-Based Raw DNA Interpretation Services: Using Social Media to Target Hard-to-Reach Populations*
*J Med Internet Res 2019;21(7):e12980*
*URL: http://www.jmir.org/2019/7/e12980/*
*doi:10.2196/12980*
*PMID:31364607*

Original Paper

# Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review

Emily G Lattie[1], PhD; Elizabeth C Adkins[1], MA; Nathan Winquist[1], BA; Colleen Stiles-Shields[2], PhD; Q Eileen Wafford[3], MST, MLiS; Andrea K Graham[1], PhD

[1]Center for Behavioral Intervention Technologies, Northwestern University, Chicago, IL, United States
[2]Department of Psychiatry and Behavioral Sciences, Rush University Medical Center, Chicago, IL, United States
[3]Galter Health Sciences Library and Learning Center, Northwestern University, Chicago, IL, United States

**Corresponding Author:**
Emily G Lattie, PhD
Center for Behavioral Intervention Technologies
Northwestern University
750 N Lake Shore Drive
10th Floor
Chicago, IL, 60611
United States
Phone: 1 3125033741
Email: emily.lattie@northwestern.edu

## Abstract

**Background:** College students are increasingly reporting common mental health problems, such as depression and anxiety, and they frequently encounter barriers to seeking traditional mental health treatments. Digital mental health interventions, such as those delivered via the Web and apps, offer the potential to improve access to mental health treatment.

**Objective:** This study aimed to review the literature on digital mental health interventions focused on depression, anxiety, and enhancement of psychological well-being among samples of college students to identify the effectiveness, usability, acceptability, uptake, and adoption of such programs.

**Methods:** We conducted a systematic review using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (registration number CRD42018092800), and the search strategy was conducted by a medical research librarian in the following databases: MEDLINE (Ovid), EMBASE (Elsevier), PsycINFO (EbscoHost), the Cochrane Library (Wiley), and Web of Science (Thomson Reuters) from the date of inception to April 2019. Data were synthesized using a systematic narrative synthesis framework, and formal quality assessments were conducted to address the risk of bias.

**Results:** A total of 89 studies met the inclusion criteria. The majority of interventions (71/89, 80%) were delivered via a website, and the most common intervention was internet-based cognitive behavioral therapy (28, 31%). Many programs (33, 37%) featured human support in the form of coaching. The majority of programs were either effective (42, 47%) or partially effective (30, 34%) in producing beneficial changes in the main psychological outcome variables. Approximately half of the studies (45, 51%) did not present any usability or acceptability outcomes, and few studies (4, 4%) examined a broad implementation of digital mental health interventions on college campuses. Quality assessments revealed a moderate-to-severe risk of bias in many of the studies.

**Conclusions:** Results suggest that digital mental health interventions can be effective for improving depression, anxiety, and psychological well-being among college students, but more rigorous studies are needed to ascertain the effective elements of these interventions. Continued research on improving the user experience of, and thus user engagement with, these programs appears vital for the sustainable implementation of digital mental health interventions on college campuses.

***(J Med Internet Res 2019;21(7):e12869)*** doi:10.2196/12869

**KEYWORDS**

eHealth; mHealth; mental health; students; universities

## Introduction

In the last decade, rising rates of college students experiencing symptoms of depression and anxiety have been observed [1-3]. Globally, approximately 31% of college students screened positive for a mental health disorder over the course of the last year [4]. It has also been increasingly recognized that accessing treatment for these common mental health problems is difficult. Many students have low mental health literacy and do not recognize a need for treatment but rather believe that these depression and anxiety symptoms are typical college stress and, therefore, do not warrant treatment [5]. Students who do recognize a need for mental health services often face multiple barriers to accessing care, perceive the care available to them as inconvenient, and are skeptical about the efficacy of care [6,7].

Campus counseling centers are well positioned to provide mental health care. However, many counseling centers across the country are underresourced, have difficulty reaching students in need, and operate at full capacity during much of the year [8].

Digital mental health interventions, such as those delivered via mobile- and Web-based platforms, offer the possibility of treatment to college students with common mental health problems while circumventing many existing barriers to receiving traditional mental health services, including stigma and time [9-11].

The evidence base for digital mental health interventions for general adult populations is vast [12-15], and the evidence base for college and university student populations is rapidly accumulating. In 2013, a systematic review of technology-based interventions for mental health in tertiary students found that these types of interventions offer promise for improving symptoms of certain mental health problems, but it concluded that more research was needed [16]. A 2014 systematic review and meta-analysis of computer-delivered and Web-based interventions for university students found that these types of interventions can be effective in improving depression, anxiety, and stress among students [17]. More recently, a 2018 systematic review and meta-analysis found that internet interventions can have small-to-moderate effects on a range of mental health conditions [18].

However, there have been limitations of these past reviews, as they have focused exclusively on studies that were randomized controlled trials (RCTs). Although a focus on studies with RCT designs allows researchers to evaluate the efficacy and effectiveness of digital mental health interventions, the exclusion of papers reporting on other study designs presents a significant gap in our ability to assess the uptake and adoption of digital mental health interventions for university students (which could be assessed in nonrandomized designs, including single-arm trials in which an intervention is made available to all students on campus). This is particularly important as the full public health potential of these types of interventions is tied not only to clinical efficacy but also to the successful implementation of these programs in real-world settings. Across the board, the implementation and integration of digital health tools into routine care settings has been a challenge. Many have called for testing digital health tools under more pragmatic conditions to maximize the transfer of knowledge from research trials to real-world implementation [19-21], and studies examining the real-world uptake and engagement with digital mental health tools have generally found low engagement and completion rates [22]. Furthermore, in recent years, increased focus has been on assessing the user experience (including the usability and acceptability) of such interventions to identify and rectify user experience failings that could limit one's ability and desire to continue to use a program [23-25]. The aim of this systematic review was to evaluate the effectiveness, usability, acceptability, uptake, and adoption of digital mental health interventions for treating depression and anxiety and for enhancing psychological well-being among college students. Characteristics of the student digital mental health interventions have been described here.

## Methods

### Eligibility Criteria

To be included in this review, studies had to (1) examine an intervention that aimed to improve psychological well-being, psychological distress, stress, depressive, and/or anxious symptoms; (2) deliver the intervention via a digital platform (including mobile phone, website, virtual reality systems, and offline computer programs; they could be delivered as an adjunct to face-to-face interventions); (3) include students enrolled in higher education institutions, such as 2-year community colleges, professional schools (eg, medical school and nursing school), 4-year colleges (ie, bachelor's degree–granting institutions that do not offer graduate degrees), and universities; (4) report outcomes related to psychological well-being, psychological distress, stress, depressive and anxious symptoms, and/or the use and reach of an intervention; and (5) be written in English. In this paper, we use the term *college students* to refer to all students in postsecondary education, including medical students. All study designs were included, with the exception of technical validation papers reporting exclusively on the development of digital mental health interventions. Conference abstracts were also excluded.

### Search Strategy

A comprehensive search strategy was developed using keywords and controlled vocabulary to describe university students, depression and anxiety, and digital mental health interventions. The search strategy was adapted to the electronic databases MEDLINE (Ovid), EMBASE (Elsevier), PsycINFO (EBSCOhost), Web of Science (Thomson Reuters), and the Cochrane Library (Wiley). Each database was searched from the date of inception to April 18, 2019. As some relevant journals (ie, *JMIR Mental Health* and *Digital Health*) are not indexed in the searched sources, an additional handsearch was conducted through these publications and through the reference lists of related systematic reviews. The searches were not limited based on publication date, language, document type, or study design. Throughout the study selection, the reference lists of included studies were further reviewed to identify relevant citations. The search strategy terms are presented in Multimedia Appendix 1. The review adhered to the Preferred Reporting

Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [26] and was registered before data extraction on the international prospective register of systematic reviews PROSPERO website (registration number CRD42018092800).

### Study Selection

Search results were uploaded into Rayyan, a Web-based software program that allows for reviewers to collaborate during the study selection process [27]. Two reviewers independently screened each of the titles and abstracts from the initial literature search against the inclusion criteria. Authors EGL, ECA, NW, CSS, and AKG served as reviewers. Full-text articles for the approved articles were then screened independently by 2 reviewers. Discrepancies about inclusion were resolved by discussion, and a third reviewer was brought into the discussion if necessary.

### Data Extraction

Two reviewers extracted the data independently from each eligible study using a Web-based extraction form that was piloted and calibrated with all reviewers before formal data extraction. Discrepancies about data extraction were resolved by discussion, and a third reviewer was brought into the discussion if necessary. The data extracted included the study location, study design, type of comparator, type of prevention/treatment, type of technology, name of technology/program, type of program, primary intervention target(s), presence of support, student population, setting, sample size, length of intervention, usability and acceptability outcomes, uptake and adoption outcomes, psychological outcomes, and type of analyses performed (ie, completer or intent to treat).

### Outcomes

This review examined the effectiveness, usability, acceptability, uptake, and adoption of digital mental health interventions for treating depression and anxiety and for enhancing psychological well-being among college students. The effectiveness outcomes included measures of depressive symptomatology (eg, Beck Depression Inventory-II [28] and Patient Health Questionnaire [29]), measures of anxious symptomatology (eg, Beck Anxiety Inventory [30] and Anxiety Sensitivity Inventory [31]), and measures of psychological distress and well-being (eg, Perceived Stress Scale [32] and Scales of Psychological Well-being [33]).

For the purpose of this review, usability was defined as the quality of a user's experience when interacting with a program. Usability is an umbrella term that includes the ease of learning a program, the efficiency of use, the memorability of it, and the subjective satisfaction with a program. The usability outcomes include standard usability measures (eg, the System Usability Scale [34]) and qualitative usability reports. Acceptability is specifically about satisfaction with different aspects of the program and was primarily measured through qualitative self-reporting.

For the purpose of this review, the terms *uptake* and *adoption* were used in conjunction with one another [35] and were together defined as the action of trying an innovation. Thus, uptake and adoption outcomes were primarily metrics on the number of downloads and uses and were intended to be described alongside the number of users relative to the population of potential users when available (to determine the rates of service penetration [35]). However, few studies provided these details. Metrics on the completion of follow-up assessments and study attrition (or fidelity to the intervention) were examined in studies that did not provide detailed program usage metrics to allow for further implementation-related insights to be gathered.

### Quality Assessment

As this review included both randomized trials and nonrandomized trials, the risk of bias was assessed using 2 separate tools: the Cochrane Collaboration's tool for assessing risk of bias in randomized trials [36] and the Cochrane Collaboration's tool for assessing risk in nonrandomized studies of interventions [37]. For randomized trials, risk of bias was evaluated for selection bias, performance bias, detection bias, attrition bias, and reporting bias using the anchors of a low, high, or unclear risk of bias. For nonrandomized trials, risk of bias was evaluated for bias because of confounding, bias in selection of participants into the study, bias in classification of interventions, bias because of deviations from intended interventions, bias because of missing data, bias in measurement of outcomes, and bias in selection of the reported result.

### Data Synthesis

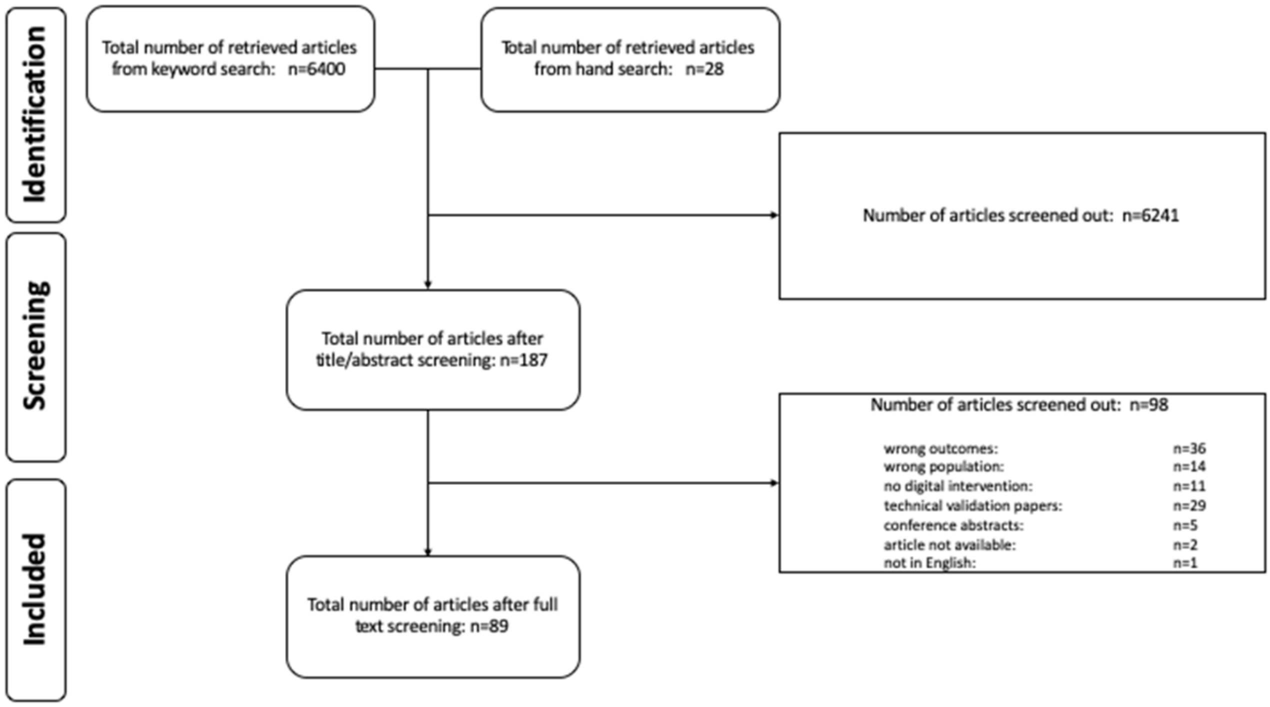
A systematic narrative framework was used to synthesize the data [38-40]. Owing to the high degree of heterogeneity in outcomes and measurement included in this study, a meta-analytic approach was not appropriate. Following the systematic narrative framework for literature reviews [39], the results of included studies were synthesized and presented without reference to the statistical significance of the findings.

## Results

### Included Studies

A total of 6428 article titles and abstracts were reviewed. Then, 187 full-text articles were reviewed for inclusion, with 89 studies included in the review for data extraction. See Figure 1 for the PRISMA flow diagram.

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




## Study Characteristics

Of the 89 studies included in this review, 72 employed randomized study designs [41-112], whereas the remaining 17 were nonrandomized [113-129]. One of the included studies [83] was a secondary analysis of another included study [84]. Another one of the included studies [129] was a qualitative study of the usability and acceptability of another included study [101]. The vast majority of the included studies (n=81 [41-44,47-52,54-56,58-111,113,114,116-120,122,123,125-129]) took place at universities, with far fewer taking place at 4-year colleges (n=2 [45,46]), at health professional schools (eg, medical school and nursing school; n=5 [57,115,121,124,130]), and at community colleges (n=1 [53]). Approximately half of the studies (n=44 [42-48,50,53,55,56,58-61,63,65,66,72,73,76-78,81,87-90,93,95,98,99,101-104,106,108,110,116-118,122,125]) targeted undergraduate students exclusively. Most studies included were focused on either universal prevention programs (n=36 [43,48,53,54,63-67,70-73,75-78,90,91,94,95,97,99,100,102,106-108,110-112,114,118,119,125,126]), or on treatment intervention programs (n=22 [41,49,50,52,58,69,74,79,80,83-86,88,89,92,98,113,117,120,122,127]).

We examined if each study appeared to be designed specifically for students (eg, the purpose of the study was focused on college student mental health) or if college students appeared to be used as a convenience sample. As seen in Multimedia Appendix 2, a minority of the studies (n=12 [44,71,89,90,93,95,104,110,111,114,118,123]) appeared to use college students as a convenience sample, and for an additional 9 studies, it was unclear if the program was being tested specifically for college student mental health or if college students were a convenience sample [47,50,51,61,64,68,84,105,122].

The majority of these studies (n=46 [43-48,51-55,57-59,62,65-67,69,72,76-78,86,87,113,116,119,120,122,124-127,131]) took place in the United States. As seen in Multimedia Appendix 2, several studies also took place in the United Kingdom (n=6 [60,73,75,80,96,114]), Ireland (n=5 [83-85,117,123]), Australia (n=5 [50,61,74,109,115]), Canada (n=5 [42,49,81,102,103]), and China (n=5 [70,71,89,91,100]).

A total of 71 studies utilized a Web-based technology ([42-44,46,49,50,53-55,57,59-88,90,91,94,96-101,103-110,112,113,115-121,123-126,128]), and 8 studies utilized interventions delivered via mobile phone (app-based programs and short message service–based programs) [52,56,59,95,101,102,114,129]. Furthermore, 11 studies utilized offline computer-based programs [41,45,47,51,52,89,92,93,105,111,127], and 3 studies focused on virtual reality programs [58,111,122]. As seen in Multimedia Appendix 2, some studies included more than one type of technology, and thus, the numbers reported in the previous sentence add up to more than the 89 studies included in this review. Although the interventions examined were variable in content and in length, the most common type of intervention examined was internet-based cognitive behavioral therapy (n=28 [42,49,50,57,61,68,72-75,79-81,83-86,88,98,101,103,107,109,112,113,123,128,129]). The modal length of intervention was 8 weeks. Most studies focused on digital mental health programs developed for the specific study, whereas a minority of studies (n=15) focused on publicly available interventions (ie, Beating the Blues, n=3 [83,84,86]; MoodGYM, n=5 ([50,57,68,98,112]); Therapist-Assisted Online, n=1 [113]; Headspace and Smiling Mind, n=1 [95]; Family eJournal, n=1 [90]; Tess chatbot, n=1 [97]; DeStressify, n=1 [102]; Overcome Social Anxiety, n=1 [103]); SilverCloud Health's Space from Depression, Space from Anxiety, and Space from Stress, n=1 [128]; and CareCollaborateConnect: Student Success and thedesk, n=1 [109]).

The majority of interventions studied offered some level of support or guidance to users—many of these interventions featured coaching from a human (n=33 [41,49-51,58,62-66,69,71,73,74,76,79,82-88,94,101,104,105,107,116,122,123,128,129]), whereas others featured automated support (n=18 [44,48,53,57,59,67,70,77,91,98-101,106,108,112,113,115]), such as prescripted emails. Several studies examined self-guided interventions (n=33 [42,43,45,46,52,54-56,60,61,68,72,75,78,80,81,89,90,92,93,95,97,102,103,109-111,114,118,119,121,124,125]) in which the participants only had contact with the study staff for research assessments; few studies focused on technology as an adjunct to therapy (n=5 [47,86,100,120,126]), 2 studies provided peer support [96,117], and 1 study had an unclear presence of support (n=1 [127]).

### Effectiveness of Interventions Studied on Psychological Outcomes

As seen in Multimedia Appendix 3, the majority of studies reported that the digital mental health intervention(s) of interest were either effective (eg, improvements were observed in all main outcomes; n=42 [41,42,45,47-51,54,56,57,59,60,62,63,68,71,78,79,84-89,92,93,100,101,103,112,113,115,120,122,123,125,127,128,131]) or partially effective (eg, improvements were observed in some main outcomes; n=30 [43,44,52,53,58,61,65,67,69,70,72-77,80-82,90,91,94-97,99,102,108, 111,124]) in producing beneficial changes in the main psychological outcome variables. A minority of studies reported on interventions that were not effective (n=10 [46,55,64,66,98,114,117]) or did not report on psychological outcomes and focused on program usage or usability (n=6 [116,118,119,121,126,129]). An examination of intervention effectiveness by type of technology used follows. Of the 71 studies that included a Web-based intervention, 30 were effective, 25 were partially effective, 8 were not effective at producing beneficial changes in the main psychological outcome variables, and for 8 studies, effectiveness was not applicable. Of the 8 studies that included interventions delivered via mobile phone, 3 were effective, 3 were partially effective, 1 was not effective, and for 1 study, effectiveness was not applicable. Of the 11 studies that included offline computer-based programs, 10 were effective and 1 was partially effective. Of the 3 studies that included a virtual reality–based intervention, 1 was effective and 2 were not effective.

Of the 42 studies deemed effective in producing beneficial changes in the main psychological outcome variables, 36 of those studies utilized a control condition; although the quality of control conditions varied broadly and ranged from treatment as usual or waitlists to other digital programs or face-to-face treatment [41,42,45,47,48,50,51,54,56,57,59,60,71,78,79,84,86,88,89,92,93,100,101,103,105,107,109,112,113,125]. See Multimedia Appendix 2 for more information. Some studies targeted general mental health and well-being and listed several primary outcome variables; thus, it was difficult to ascertain whether the intervention was *fully* effective. To provide a conservative estimate, these studies were counted as partially effective.

### Usability and Acceptability of Interventions Studied

Approximately half of the studies included in this review (n=45 [41-44,48,54-57,59,60,62,63,70,73,76-78,81,84,85,87-93,96,100,102,104,106,110,111-113,116,118,121-125,127]) did not present any usability or acceptability outcomes. In those studies that presented usability and acceptability outcomes, the results were generally favorable. However, response rates were often low (which was specifically noted in the studies by Lintvedt et al [68] and Mailey et al [69]), so it is difficult to ascertain the true acceptability of these interventions as those who continued to engage with the study procedures may have found the interventions more useful and usable than those who did not.

As seen in Multimedia Appendix 3, studies that presented usability and/or acceptability outcomes typically relied on single-item Likert scales, questionnaires, or user feedback interviews rather than validated measures. A minority of studies used validated measures such as the System Usability Scale [65-67,120,132] to assess usability and the Client Satisfaction Questionnaire [61,71,86,133] to assess acceptability. The results of formal usability testing were not presented in any of the studies.

### Uptake and Adoption of Interventions Studied

The vast majority of studies did not specify the size of the pool of potential participants from which the study participants were drawn (n=81 [41-54,56-61,63-71,73-91,93-101,103-105,107-111,113-115,117-127,129]) and recruited from seemingly large pools of students. Furthermore, many studies did not present the metrics on the usage of the digital mental health interventions (n=29 [44,45,48,50,53,56,69,70,73,74,76-78,81,89,90,93,98,99,103-105,107,109-111,113,114,124]. As detailed program usage metrics were not addressed by the authors of several included studies, the *Uptake and Adoption* column in Multimedia Appendix 3 includes additional data on completion of follow-up assessments and study attrition as a proxy for intervention uptake and adoption. For studies that did not provide clear metrics on the completion of follow-up assessments and/or study attrition, we have listed *not reported* in this column. For studies that examined digital mental health interventions in standardized laboratory settings, we have listed *N/A due to standardized within-lab use*.

Relatively few studies examined the implementation of a digital mental health intervention on a college campus and reported on the implementation outcomes [35,93,98]. Although a small handful of studies reported on the broad uptake and adoption of programs that were implemented on college campuses [86,116,119,128], the Beating the Blues implementation by Santucci et al [86] and the SilverCloud implementation by Palacios et al [128] were the only studies in which the feasibility of implementing a digital mental health program was explicitly discussed. Santucci et al set out to assess the feasibility, acceptability, and effectiveness of Beating the Blues for university students (as benchmarked to published trials) and outlined their process of conducting a needs assessment and engaging stakeholders before commencing the trial [86]. They found preliminary support for the feasibility of disseminating and implementing Beating the Blues in a university health center and effectiveness similar to what had been documented in a previous RCT. Palacios et al [128] conducted an open trial in which 3 SilverCloud programs—Space from Depression, Space from Anxiety, and Space from Stress—were made available to the students and were advertised through on-campus counseling centers. The majority of the participants found the programs helpful and found benefit in having a supported Web-based intervention available on campus, and the programs demonstrated feasibility, acceptability, and effectiveness.

As seen in Multimedia Appendix 3, many studies had high rates of attrition and low rates of sustained program use. Although usage was variable and cannot be directly compared across studies, a pattern emerged such that for module-based interventions, usage dropped over an individual's time spent in the study. For example, module 1 program completion rates were generally high, and in many studies examined, a minority of participants completed all available modules.

### Risk of Bias

As seen in Table 1, of the 72 randomized studies, 28 studies were judged as having a low risk of bias ([41,44,47,54,57,60,64,66,75,78,84,86,88,91,92,94,95,97,99,100,101,102-104,109,111,112]) and only 9 studies were judged as having a high risk of bias [45,49,58,63,65,72,83,96,110]. The remaining 35 studies were judged as having some concerns regarding bias [42,43,46,48,50-53,55,56,59,61,62,67-71,73,74,76,77,79-82,85,87,89,90,93,98,105-108,131]. Risk of bias most frequently emerged because of a potential bias in measurement of the outcome. Some concerns were noted in this domain for roughly one-third of the studies (n=27 [42,46,49,50,52,53,55,56,58,59,61-63,65,67-70,72-74,76,81,82,85,87,89]) because outcomes were self-reported in nature and the participants were aware of the intervention they received. Concerns were often noted regarding potential bias arising from the randomization process, and potential bias because of missing outcome data that were not analyzed in a manner to minimize risk [80-82]. Potential bias arising from the randomization process was frequently noted because of baseline imbalances that suggested problems with randomization. Concerns surrounding risk of bias because of missing outcome data were frequently related to high levels of attrition, which is a particularly common problem in studies of digital mental health interventions.

Of the 16 nonrandomized studies screened for risk of bias, the majority of studies (n=10 [113-115,117,120,122-125,128]) demonstrated a serious risk of bias using the Risk of Bias in Non-Randomized Studies of Interventions (ROBINS-I) rating scale, and 1 study demonstrated a critical risk of bias [127]. One qualitative study [129] from which we extracted usability and acceptability data was not included in the risk-of-bias assessments. This study collected data from participants of an already analyzed RCT [101]. The details are provided in Table 2. Risk of bias is more common in nonrandomized studies, and the 3 studies that were determined to have a low risk of bias [116,118,119] all reported on the uptake and adoption of digital mental health programs as their main outcomes. As these metrics were objective, minimal risk of bias was identified. For the majority of studies, a potential for bias was identified in the measurement of outcomes, as outcomes were self-reported in nature and the participants were aware of the intervention they received.

**Table 1.** Risk of bias for randomized studies.

| Authors and year of publication | Dmn[a] #1[b] | Dmn #2[c] | Dmn #3[d] | Dmn #4[e] | Dmn #5[f] | Overall risk |
|---|---|---|---|---|---|---|
| Alvarez et al, 2008 [41] | LR[g] | LR | LR | LR | LR | LR |
| Arpin-Cribbie et al, 2012 [42] | LR | LR | LR | SC[h] | LR | SC |
| Asbury et al, 2018 [90] | SC | SC | LR | LR | LR | SC |
| Auyeung & Mo, 2018 [91] | LR | LR | LR | LR | LR | LR |
| Bedford et al, 2018 [92] | LR | LR | LR | LR | LR | LR |
| Booker & Dunsmore, 2017 [43] | LR | SC | SC | LR | LR | SC |
| Braithwaite & Fincham, 2009 [44] | LR | LR | LR | LR | LR | LR |
| Braithwaite & Fincham, 2007 [93] | SC | LR | SC | LR | LR | SC |
| Buglione et al, 1990 [45] | SC | SC | SC | LR | LR | HR[i] |
| Chiauzzi et al, 2008 [46] | LR | LR | LR | SC | LR | SC |
| Cohen et al, 1999 [47] | LR | LR | LR | LR | LR | LR |
| Cukrowicz & Joiner, 2007 [48] | SC | LR | LR | LR | LR | SC |
| Day et al, 2013 [49] | LR | LR | HR | SC | LR | HR |
| Ellis et al, 2011 [50] | LR | LR | LR | SC | LR | LR |
| Eustis et al, 2018 [94] | LR | LR | LR | LR | LR | LR |
| Fernandez et al, 1986 [51] | LR | LR | LR | LR | SC | SC |
| Fitzpatrick et al, 2017 [52] | LR | LR | LR | SC | LR | SC |
| Flett et al, 2019 [95] | LR | LR | LR | LR | LR | LR |
| Frazier et al, 2015 [53] | LR | LR | LR | SC | LR | SC |
| Freeman et al, 2008 [96] | SC | LR | HR | HR | LR | HR |
| Frith & Loprinzi, 2017 [54] | LR | LR | LR | LR | LR | LR |
| Fulmer et al, 2018 [97] | LR | LR | LR | LR | LR | LR |
| Geisner et al, 2015 [55] | LR | LR | LR | SC | LR | LR |
| Gibbel, 2010 [98] | LR | SC | LR | LR | LR | SC |
| Grassi et al, 2011 [56] | LR | LR | SC | SC | LR | SC |
| Greer, 2015 [99] | LR | LR | LR | LR | LR | LR |
| Guille et al, 2015 [57] | LR | LR | LR | LR | LR | LR |
| Hall et al, 2018 [100] | LR | LR | LR | LR | LR | LR |
| Harrer et al, 2018 [101] | LR | LR | LR | LR | LR | LR |
| Harris et al, 2002 [58] | HR | LR | HR | SC | LR | HR |
| Hintz et al, 2015 [59] | LR | LR | LR | SC | LR | SC |
| Hoppitt et al, 2014 [60] | LR | LR | LR | LR | LR | LR |
| Howell et al, 2018 [112] | LR | LR | LR | LR | LR | LR |
| Kenardy et al, 2003 [61] | LR | LR | LR | SC | LR | SC |
| King et al, 2015 [62] | LR | LR | LR | SC | LR | SC |
| Koydemir & Sun-Selisik, 2016 [63] | SC | SC | HR | SC | LR | HR |
| Kvillemo et al, 2016 [64] | LR | LR | LR | LR | LR | LR |
| Lee & Jung, 2018 [102] | LR | LR | LR | LR | LR | LR |
| Levin et al, 2014 [65] | LR | LR | LR | SC | HR | HR |
| Levin et al, 2016 [66] | LR | LR | LR | LR | LR | LR |
| Levin et al, 2017 [67] | LR | LR | LR | SC | LR | SC |

| Authors and year of publication | Dmn[a] #1[b] | Dmn #2[c] | Dmn #3[d] | Dmn #4[e] | Dmn #5[f] | Overall risk |
|---|---|---|---|---|---|---|
| Lintvedt et al, 2013 [68] | SC | LR | LR | SC | LR | SC |
| Mailey et al, 2010 [69] | LR | LR | LR | SC | LR | SC |
| Mak et al, 2015 [70] | LR | LR | LR | SC | LR | SC |
| Mak et al, 2017 [71] | LR | LR | LR | LR | SC | SC |
| McCall et al, 2018 [103] | LR | LR | LR | LR | LR | LR |
| Melnyk et al, 2015 [72] | SC | LR | LR | SC | SC | HR |
| Mogoaşe, 2013 [104] | LR | LR | LR | LR | LR | LR |
| Morris et al, 2016 [73] | LR | LR | LR | SC | LR | SC |
| Mullin et al, 2015 [74] | SC | LR | LR | SC | LR | SC |
| Musiat et al, 2014 [75] | LR | LR | LR | LR | LR | LR |
| Nguyen-Feng et al, 2015 [76] | LR | LR | LR | SC | LR | SC |
| Nguyen-Feng et al, 2016 [77] | LR | LR | LR | LR | SC | SC |
| Nguyen-Feng et al, 2017 [78] | LR | LR | LR | LR | LR | LR |
| Nordmo et al, 2015 [79] | SC | LR | LR | LR | LR | SC |
| Orbach et al, 2007 [80] | LR | LR | SC | LR | SC | SC |
| Radhu et al, 2012 [81] | LR | LR | LR | SC | LR | SC |
| Rasanen et al, 2016 [82] | LR | LR | LR | SC | LR | SC |
| Richards & Timulak, 2013 [83] | LR | SC | SC | LR | HR | HR |
| Richards et al, 2013 [84] | LR | LR | LR | LR | LR | LR |
| Richards et al, 2016 [85] | LR | LR | LR | SC | LR | SC |
| Rose et al, 2013 [105] | SC | LR | LR | LR | LR | SC |
| Sagon et al, 2018 [106] | SC | LR | LR | LR | LR | SC |
| Saleh et al, 2018 [107] | SC | LR | LR | LR | LR | SC |
| Santucci et al, 2014 [86] | LR | LR | LR | LR | LR | LR |
| Sarniak, 2009 [108] | SC | SC | LR | LR | LR | SC |
| Seligman et al, 2007 [87] | SC | LR | LR | SC | LR | SC |
| Stallman et al, 2018 [109] | LR | LR | LR | LR | LR | LR |
| Taitz, 2011 [110] | SC | SC | SC | LR | LR | HR |
| Tillfors et al, 2008 [88] | LR | LR | LR | LR | LR | LR |
| Villani & Riva, 2008 [111] | LR | LR | LR | LR | LR | LR |
| Yang et al, 2015 [89] | LR | LR | LR | SC | LR | SC |

[a]Dmn: domain.
[b]Bias arising from the randomization process.
[c]Bias due to deviations from intended interventions.
[d]Bias due to missing outcome data.
[e]Bias in measurement of the outcome.
[f]Bias in selection of the reported result.
[g]LR: low risk.
[h]SC: some concerns.
[i]HR: high risk.

**Table 2.** Risk of bias for nonrandomized studies.

| Authors and year of publication | Dmn[a] #1[b] | Dmn #2[c] | Dmn #3[d] | Dmn #4[e] | Dmn #5[f] | Dmn #6[g] | Dmn #7[h] | Overall risk |
|---|---|---|---|---|---|---|---|---|
| Benton et al, 2016 [113] | MR[i] | LR[j] | SR[k] | LR | SR | SR | LR | SR |
| Carey et al, 2016 [114] | MR | LR | LR | LR | LR | SR | LR | SR |
| Finlay-Jones et al, 2016 [115] | MR | LR | LR | LR | MR | SR | LR | SR |
| Haas et al, 2008 [116] | LR | N/A[l] | N/A | N/A | N/A | N/A | N/A | LR |
| Horgan et al, 2013 [117] | MR | LR | LR | LR | SR | SR | LR | SR |
| Kaczmarek et al, 2013 [118] | LR | N/A | N/A | N/A | N/A | N/A | N/A | LR |
| Kim et al, 2011 [119] | LR | N/A | N/A | N/A | N/A | N/A | N/A | LR |
| Levin et al, 2015 [120] | MR | LR | LR | LR | MR | SR | LR | SR |
| Moir et al, 2015 [121] | MR | LR | SR | LR | NI[m] | SR | LR | NI |
| North et al, 2002 [122] | MR | LR | LR | LR | LR | SR | LR | SR |
| Palacios et al, 2018 [128] | MR | LR | LR | LR | MR | SR | LR | SR |
| Sharry et al, 2013 [123] | MR | LR | LR | LR | MR | SR | LR | SR |
| Spadaro & Hunker, 2016 [124] | MR | LR | LR | LR | LR | SR | LR | SR |
| Trockel et al, 2011 [125] | CR[n] | LR | LR | LR | SR | SR | LR | SR |
| Williams et al, 2014 [126] | MR | LR | LR | LR | LR | LR | LR | MR |
| Wilson et al, 1991 [127] | MR | CR | LR | LR | LR | SR | CR | CR |

[a]Dmn: domain.
[b]Bias due to confounding.
[c]Bias in selection of participants into the study.
[d]Bias in classification of interventions.
[e]Bias due to deviations from intended interventions.
[f]Bias due to missing data.
[g]Bias in measurement of outcomes.
[h]Bias in selection of the reported result.
[i]MR: moderate risk.
[j]LR: low risk.
[k]SR: serious risk.
[l]N/A: not applicable. Any study deemed low risk in Domain #1 is considered low risk as a whole; thus, other domains are N/A.
[m]NI: no information.
[n]CR: critical risk.

## Discussion

### Principal Findings

This study aimed to synthesize the literature on the effectiveness, usability, acceptability, uptake, and adoption of digital mental health interventions for (1) treating depression and anxiety and (2) enhancing psychological well-being among college students. In doing so, the types of interventions that have been developed and tested were characterized. The vast majority of included studies reported that the digital mental health interventions of interest were either effective, or partially effective, in producing beneficial changes in the main psychological outcome variables. This is consistent with past meta-analyses on digital mental health programs for college students [17,18] and is consistent with the broader literature on digital mental health interventions [134,135]. Effectiveness did not appear to substantially vary by type of digital mental health intervention, indicating that computer-, Web-, mobile-, and virtual reality–based interventions all hold potential for improving mental health on college campuses.

The majority of programs were studied on university campuses and enrolled broad samples of undergraduate and graduate students. The focus on universities was not surprising, as many studies were conducted at the university with which the researchers were affiliated, likely because of a combination of ease and investment in one's own community. Fewer studies took place within health professional (eg, medical school and nursing school) programs. It was notable that only 1 study comprised a community college sample, as it is widely recognized that community college students have higher rates of unmet mental health needs compared with students in traditional 4-year colleges and universities [136-138]. Furthermore, community college students are likely to face additional barriers to accessing care as they are more likely to attend school part-time while balancing other responsibilities and commitments, and many community college campuses do not provide mental health services [138]. Thus, this appears to be a priority area for further research and intervention development.

College students are often used as a convenience sample for psychological research [139]. Therefore, to interpret the findings in light of whether the included studies aimed to specifically target students, we examined whether the designs appeared specific to college student mental health. We found that the majority of studies (n=68) were focused explicitly on college student mental health, as opposed to using students for convenience sampling. This majority finding highlights the potential for these programs to be more broadly disseminated and implemented on college campuses.

Similar to what has been observed in digital mental health intervention programs for general adult populations [22,134], there were notable rates of participant attrition and early program discontinuation in many of the studies. An individual may discontinue use of a digital mental health intervention for a variety of factors. These include positive reasons, such as early mood improvements resulting in the individual no longer having a need for the intervention tools. More often though, early discontinuation of such programs appears be the result of an unsatisfying user experience. Although user experience is multifaceted, the core components of user experiences include the program's usability and acceptability, which were a focus of this study. A recent review of user engagement with mental health apps found that problems emerge because apps (1) are not designed with users in mind, (2) do not address problems users care most about, (3) do not respect user privacy, (4) are not seen as trustworthy, and (5) are unhelpful in emergencies [140]. Although this review focused on mental health apps [140], these themes appear to be translatable to reasons for poor engagement with other types of digital mental health interventions. Indeed, these themes suggest usability problems, which decrease the likelihood of user engagement because of a mismatch of design with user needs [141,142]. The majority of interventions included in this review were unnamed programs developed for research purposes, and although many reported on participant satisfaction with the program, the extent to which the interventions were tested for user experience before these trials remains largely unknown. Utilization of user-centered design and usability testing is a growing practice in digital health interventions for depression [143], smoking cessation [144,145], and diverse patient groups [146]. Indeed, this practice promotes the likelihood that the intervention is appropriately engaging, intuitive to use, and pleasing to the intended user population. Therefore, assessing for usability is a critical component in establishing the feasibility, efficacy, and generalizability of a digital health intervention, particularly for specialty populations [147].

Although user-centered design can produce programs that are more engaging and enjoyable for users, design principles alone are unlikely to produce interventions that are sustainably used on college campuses. The research-to-practice gap for digital mental health interventions is increasingly being recognized, and leaders in the field have proposed strategies to routinely incorporate implementation science methods into the study of digital mental health interventions [148,149]. These models highlight the importance of the systems in which interventions are to be introduced, and although all studies included in this review focused on college students as a population of interest, very few examined college campus systems and tested the implementation of programs onto campuses [86,109,116,119]. Increasingly, calls are being made to collect implementation-relevant data while testing new digital mental health programs or testing existing programs on new populations [150-152]. For digital mental health to fully realize its potential for college students, digital health researchers need to embrace methods and models from implementation science, such as hybrid trial designs [153], and add to the body of knowledge on how to create and support a campus mental health system that actively uses digital mental health programs.

### Strengths and Limitations

This study should be interpreted in light of its strengths and limitations. Consistent with best practices, the articles were reviewed by 2 independent reviewers and risk of bias was assessed. The moderate-to-severe risk of bias found in many of the included randomized and nonrandomized trials indicates that the results reported may be biased in favor of the digital mental health tools and should be evaluated in that context. Bias primarily emerged because the outcomes were self-reported in nature and the participants were aware of the intervention they received—2 issues that are exceedingly common in digital health research. Although the search strategy was developed with an experienced research librarian and an additional handsearch was used, it is possible that some relevant publications were missed in the search. Several reviewed studies used active controls or comparison interventions that produced similar effects to the intervention of interest, so we were unable to evaluate the effectiveness of intervention ingredients to inform what components (eg, features or techniques) are relevant for achieving behavior change. Without the gold standard interventions in digital health for college students that could serve as comparisons with newly developed interventions, several studies that were reviewed used active controls or comparison interventions that produced similar effects to the intervention of interest. In addition, none of the included studies utilized noninferiority analyses. Therefore, the true efficacy of most of the interventions was unclear.

Another strength is that we did not limit this review to RCTs of computer- and Web-based programs. As such, this study expands on past work by offering a much broader look at the types of digital mental health programs that have been available for students and a look at the uptake and adoption of such interventions. Uptake and adoption could not have been meaningfully examined if this review was limited to RCTs. However, the consequence of including multiple trial designs precluded us conducting a meta-analysis because of the heterogeneity of the data included.

### Conclusions

Digital mental health interventions for depression, anxiety, and the enhancement of psychological well-being have the potential to improve the mental health of college students around the world. The majority of interventions have focused on Web-based technologies, and there remains a need for further research on interventions delivered via mobile phones. To date, published studies on digital mental health programs have primarily been focused on establishing efficacy and/or effectiveness rather than on supporting program uptake and adoption across campus communities. For these programs to realize their potential, they need to be successfully and sustainably implemented on college campuses as part of the array of available mental health services. Further research on digital mental health interventions for college students should focus on designing and testing programs that are viewed as usable and acceptable to students and on methods of implementing such programs on college campuses.

## Acknowledgments

EGL is supported by a research grant K08 MH112878 from the National Institute of Mental Health. AKG is supported by a research grant K01 DK116925 from the National Institute of Diabetes and Digestive and Kidney Diseases.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Search strategy overview.

[PDF File (Adobe PDF File), 103KB - jmir_v21i7e12869_app1.pdf ]

## Multimedia Appendix 2

Study details.

[XLSX File (Microsoft Excel File), 21KB - jmir_v21i7e12869_app2.xlsx ]

## Multimedia Appendix 3

Study results.

[XLSX File (Microsoft Excel File), 37KB - jmir_v21i7e12869_app3.xlsx ]

## References

1. Pinder-Amaker S, Bell C. A bioecological systems approach for navigating the college mental health crisis. Harv Rev Psychiatry 2012;20(4):174-188. [doi: 10.3109/10673229.2012.712842] [Medline: 22894727]
2. Eisenberg D, Hunt J, Speer N. Help seeking for mental health on college campuses: review of evidence and next steps for research and practice. Harv Rev Psychiatry 2012;20(4):222-232. [doi: 10.3109/10673229.2012.712839] [Medline: 22894731]
3. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by US college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019 Jan 1;70(1):60-63. [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
4. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO world mental health surveys international college student project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638. [doi: 10.1037/abn0000362] [Medline: 30211576]
5. Eisenberg D, Golberstein E, Gollust SE. Help-seeking and access to mental health care in a university student population. Med Care 2007 Jul;45(7):594-601. [doi: 10.1097/MLR.0b013e31803bb4c1] [Medline: 17571007]
6. Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011 May;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]
7. Mowbray CT, Megivern D, Mandiberg JM, Strauss S, Stein CH, Collins K, et al. Campus mental health services: recommendations for change. Am J Orthopsychiatry 2006 Apr;76(2):226-237. [doi: 10.1037/0002-9432.76.2.226] [Medline: 16719642]
8. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017 Nov;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
9. 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]
10. Rai M, Vigod SN, Hensel JM. Barriers to office-based mental health care and interest in e-communication with providers: a survey study. JMIR Ment Health 2016 Aug 1;3(3):e35 [FREE Full text] [doi: 10.2196/mental.6068] [Medline: 27480108]

11. Kass AE, Balantekin KN, Fitzsimmons-Craft EE, Jacobi C, Wilfley DE, Taylor CB. The economic case for digital interventions for eating disorders among United States college students. Int J Eat Disord 2017 Mar;50(3):250-258 [FREE Full text] [doi: 10.1002/eat.22680] [Medline: 28152203]
12. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
13. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
14. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
15. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Dec 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
16. Farrer L, Gulliver A, Chan JK, Batterham PJ, Reynolds J, Calear A, et al. Technology-based interventions for mental health in tertiary students: systematic review. J Med Internet Res 2013 May 27;15(5):e101 [FREE Full text] [doi: 10.2196/jmir.2639] [Medline: 23711740]
17. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
18. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759. [doi: 10.1002/mpr.1759] [Medline: 30585363]
19. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
20. Mohr DC, Cheung K, Schueller SM, Hendricks BC, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
21. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 9;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
22. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Dec 6;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
23. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;46(s2):49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
24. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
25. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
28. Beck AT, Steer RA, Brown GK. BDI-II, Beck Depression Inventory: Manual. San Antonio, Texas: The Psychological Corporation; 1996.
29. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
30. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988 Dec;56(6):893-897. [doi: 10.1037//0022-006X.56.6.893] [Medline: 3204199]
31. Peterson RA, Reiss S. Anxiety Sensitivity Index Manual. Worthington, Ohio: International Diagnostic Systems; 1992.
32. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
33. Ryff CD, Keyes CL. The structure of psychological well-being revisited. J Pers Soc Psychol 1995 Oct;69(4):719-727. [doi: 10.1037/0022-3514.69.4.719] [Medline: 7473027]
34. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, Colorado: Measuring Usability LLC; 2011.

35. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
36. Higgins JP, Sterne JA, Savović J, Page MJ, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. In: Chandler J, McKenzie J, Boutron I, Welch V, editors. Cochrane Methods 2016. New Jersey: Wiley; 2016:29.
37. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
38. Snilstveit B, Oliver S, Vojtkova M. Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. J Dev Econ 2012 Sep 27;4(3):409-429 [FREE Full text] [doi: 10.1080/19439342.2012.710641]
39. Siddaway AP, Wood AM, Hedges LV. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. Annu Rev Psychol 2019 Dec 4;70:747-770. [doi: 10.1146/annurev-psych-010418-102803] [Medline: 30089228]
40. Baumeister RF, Leary MR. Writing narrative literature reviews. Rev Gen Psychol 1997;1(3):311-320. [doi: 10.1037//1089-2680.1.3.311]
41. Alvarez LM, Cortés SJ, León SO, Estrella J, Sánchez SJ. Computer program in the treatment for major depression and cognitive impairment in university students. Comput Human Behav 2008 May;24(3):816-826. [doi: 10.1016/j.chb.2007.02.013]
42. Arpin-Cribbie C, Irvine J, Ritvo P. Web-based cognitive-behavioral therapy for perfectionism: a randomized controlled trial. Psychother Res 2012;22(2):194-207. [doi: 10.1080/10503307.2011.637242] [Medline: 22122217]
43. Booker JA, Dunsmore JC. Expressive writing and well-being during the transition to college: comparison of emotion-disclosing and gratitude-focused writing. J Soc Clin Psychol 2017 Sep;36(7):580-606. [doi: 10.1521/jscp.2017.36.7.580]
44. Braithwaite SR, Fincham FD. A randomized clinical trial of a computer based preventive intervention: replication and extension of ePREP. J Fam Psychol 2009 Feb;23(1):32-38. [doi: 10.1037/a0014061] [Medline: 19203157]
45. Buglione SA, DeVito AJ, Mulloy JM. Traditional group therapy and computer-administered treatment for test anxiety. Anxiety Stress Coping 1990 Jun;3(1):33-39. [doi: 10.1080/08917779008248739]
46. Chiauzzi E, Brevard J, Thum C, Thurn C, Decembrele S, Lord S. MyStudentBody-Stress: an online stress management intervention for college students. J Health Commun 2008 Sep;13(6):555-572. [doi: 10.1080/10810730802281668] [Medline: 18726812]
47. Cohen GE, Kerr BA. Computer-mediated counseling: an empirical study of a new mental health treatment. Comput Human Behav 1999 Apr 22;15(4):13-26. [doi: 10.1300/J407v15n04_02]
48. Cukrowicz KC, Joiner TE. Computer-based intervention for anxious and depressive symptoms in a non-clinical population. Cogn Ther Res 2007 Apr 4;31(5):677-693. [doi: 10.1007/s10608-006-9094-x]
49. Day V, McGrath PJ, Wojtowicz M. Internet-based guided self-help for university students with anxiety, depression and stress: a randomized controlled clinical trial. Behav Res Ther 2013 Jul;51(7):344-351. [doi: 10.1016/j.brat.2013.03.003] [Medline: 23639300]
50. Ellis LA, Campbell AJ, Sethi S, O'Dea BM. Comparative randomized trial of an online cognitive-behavioral therapy program and an online support group for depression and anxiety. J Cyber Ther Rehabil 2011;4(4):461-467 [FREE Full text]
51. Fernandez E, Brechtel M, Mercer A. Personal and simulated computer-aided counseling: perceived versus measured counseling outcomes for college students. J Coll Stud Dev 1986;27(3):224-228 [FREE Full text]
52. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (WOEBOT): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
53. Frazier P, Meredith L, Greer C, Paulsen JA, Howard K, Dietz LR, et al. Randomized controlled trial evaluating the effectiveness of a web-based stress management program among community college students. Anxiety Stress Coping 2015;28(5):576-586. [doi: 10.1080/10615806.2014.987666] [Medline: 25420030]
54. Frith E, Loprinzi P. Can Facebook reduce perceived anxiety among college students? Randomized controlled exercise trial using the transtheoretical model of behavior change. JMIR Ment Health 2017 Dec 8;4(4):e50 [FREE Full text] [doi: 10.2196/mental.8086] [Medline: 29222077]
55. Geisner IM, Varvil-Weld L, Mittmann AJ, Mallett K, Turrisi R. Brief web-based intervention for college students with comorbid risky alcohol use and depressed mood: does it work and for whom? Addict Behav 2015 Mar;42:36-43 [FREE Full text] [doi: 10.1016/j.addbeh.2014.10.030] [Medline: 25462652]
56. Grassi A, Gaggioli A, Riva G. New technologies to manage exam anxiety. Stud Health Technol Inform 2011;167:57-62. [Medline: 21685642]

57. Guille C, Zhao Z, Krystal J, Nichols B, Brady K, Sen S. Web-based cognitive behavioral therapy intervention for the prevention of suicidal ideation in medical interns: a randomized clinical trial. JAMA Psychiatry 2015 Dec;72(12):1192-1198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1880] [Medline: 26535958]
58. Harris SR, Kemmerling RL, North MM. Brief virtual reality therapy for public speaking anxiety. Cyberpsychol Behav 2002 Dec;5(6):543-550. [doi: 10.1089/109493102321018187] [Medline: 12556117]
59. Hintz S, Frazier PA, Meredith L. Evaluating an online stress management intervention for college students. J Couns Psychol 2015 Apr;62(2):137-147. [doi: 10.1037/cou0000014] [Medline: 24635586]
60. Hoppitt L, Illingworth JL, MacLeod C, Hampshire A, Dunn BD, Mackintosh B. Modifying social anxiety related to a real-life stressor using online cognitive bias modification for interpretation. Behav Res Ther 2014 Jan;52:45-52 [FREE Full text] [doi: 10.1016/j.brat.2013.10.008] [Medline: 24317281]
61. Kenardy J, McCafferty K, Rosa V. Internet-delivered indicated prevention for anxiety disorders: a randomized controlled trial. Behav Cognit Psychother 2003;31(3):279-289. [doi: 10.1017/S1352465803003047]
62. King CA, Eisenberg D, Zheng K, Czyz E, Kramer A, Horwitz A, et al. Online suicide risk screening and intervention with college students: a pilot randomized controlled trial. J Consult Clin Psychol 2015 Jun;83(3):630-636 [FREE Full text] [doi: 10.1037/a0038805] [Medline: 25688811]
63. Koydemir S, Sun-Selışık ZE. Well-being on campus: testing the effectiveness of an online strengths-based intervention for first year college students. Br J Guid Counc 2015 Nov 16;44(4):434-446. [doi: 10.1080/03069885.2015.1110562]
64. Kvillemo P, Brandberg Y, Bränström R. Feasibility and outcomes of an internet-based mindfulness training program: a pilot randomized controlled trial. JMIR Ment Health 2016 Jul 22;3(3):e33 [FREE Full text] [doi: 10.2196/mental.5457] [Medline: 27450466]
65. Levin ME, Pistorello J, Seeley JR, Hayes SC. Feasibility of a prototype web-based acceptance and commitment therapy prevention program for college students. J Am Coll Health 2014;62(1):20-30 [FREE Full text] [doi: 10.1080/07448481.2013.843533] [Medline: 24313693]
66. Levin ME, Hayes SC, Pistorello J, Seeley JR. Web-based self-help for preventing mental health problems in universities: comparing acceptance and commitment training to mental health education. J Clin Psychol 2016 Mar;72(3):207-225. [doi: 10.1002/jclp.22254] [Medline: 26784010]
67. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017 Dec 1;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
68. Lintvedt OK, Griffiths KM, Sørensen K, Østvik AR, Wang C, Eisemann M, et al. Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: a randomized controlled trial. Clin Psychol Psychother 2013;20(1):10-27. [doi: 10.1002/cpp.770] [Medline: 21887811]
69. Mailey EL, Wójcicki TR, Motl RW, Hu L, Strauser DR, Collins KD, et al. Internet-delivered physical activity intervention for college students with mental health disorders: a randomized pilot trial. Psychol Health Med 2010 Dec;15(6):646-659. [doi: 10.1080/13548506.2010.498894] [Medline: 21154018]
70. Mak WW, Chan AT, Cheung EY, Lin CL, Ngai KC. Enhancing web-based mindfulness training for mental health promotion with the health action process approach: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e8 [FREE Full text] [doi: 10.2196/jmir.3746] [Medline: 25599904]
71. Mak WW, Chio FH, Chan AT, Lui WW, Wu EK. The efficacy of internet-based mindfulness training and cognitive-behavioral training with telephone support in the enhancement of mental health among college students and young working adults: randomized controlled trial. J Med Internet Res 2017 Dec 22;19(3):e84 [FREE Full text] [doi: 10.2196/jmir.6737] [Medline: 28330831]
72. Melnyk BM, Amaya M, Szalacha LA, Hoying J, Taylor T, Bowersox K. Feasibility, acceptability, and preliminary effects of the COPE online cognitive-behavioral skill-building program on mental health outcomes and academic performance in freshmen college students: a randomized controlled pilot study. J Child Adolesc Psychiatr Nurs 2015 Aug;28(3):147-154. [doi: 10.1111/jcap.12119] [Medline: 26268362]
73. Morris J, Firkins A, Millings A, Mohr C, Redford P, Rowe A. Internet-delivered cognitive behavior therapy for anxiety and insomnia in a higher education context. Anxiety Stress Coping 2016 Dec;29(4):415-431. [doi: 10.1080/10615806.2015.1058924] [Medline: 26079158]
74. Mullin A, Dear BF, Karin E, Wootton BM, Staples LG, Johnston L, et al. The UniWellbeing course: a randomised controlled trial of a transdiagnostic internet-delivered cognitive behavioural therapy (CBT) programme for university students with symptoms of anxiety and depression. Internet Interventions 2015 May;2(2):128-136 [FREE Full text] [doi: 10.1016/j.invent.2015.02.002]
75. Musiat P, Conrod P, Treasure J, Tylee A, Williams C, Schmidt U. Targeted prevention of common mental health disorders in university students: randomised controlled trial of a transdiagnostic trait-focused web-based intervention. PLoS One 2014;9(4):e93621 [FREE Full text] [doi: 10.1371/journal.pone.0093621] [Medline: 24736388]
76. Nguyen-Feng VN, Frazier PA, Greer CS, Howard KG, Paulsen JA, Meredith L, et al. A randomized controlled trial of a web-based intervention to reduce distress among students with a history of interpersonal violence. Psychol Violence 2015;5(4):444-454. [doi: 10.1037/a0039596]

77. Nguyen-Feng VN, Frazier PA, Greer CS, Meredith L, Howard K, Paulsen J. Testing the efficacy of three brief web-based interventions for reducing distress among interpersonal violence survivors. Transl Issues Psychol Sci 2016;2(4):439-448. [doi: 10.1037/tps0000099]
78. Nguyen-Feng VN, Greer CS, Frazier P. Using online interventions to deliver college student mental health resources: evidence from randomized clinical trials. Psychol Serv 2017 Nov;14(4):481-489. [doi: 10.1037/ser0000154] [Medline: 29120206]
79. Nordmo M, Sinding AI, Carlbring P, Andersson G, Havik OE, Nordgreen T. Internet-delivered cognitive behavioural therapy with and without an initial face-to-face psychoeducation session for social anxiety disorder: a pilot randomized controlled trial. Internet Interv 2015 Nov;2(4):429-436. [doi: 10.1016/j.invent.2015.10.003]
80. Orbach G, Lindsay S, Grey S. A randomised placebo-controlled trial of a self-help internet-based intervention for test anxiety. Behav Res Ther 2007 Mar;45(3):483-496. [doi: 10.1016/j.brat.2006.04.002] [Medline: 16814744]
81. Radhu N, Daskalakis ZJ, Arpin-Cribbie CA, Irvine J, Ritvo P. Evaluating a web-based cognitive-behavioral therapy for maladaptive perfectionism in university students. J Am Coll Health 2012;60(5):357-366. [doi: 10.1080/07448481.2011.630703] [Medline: 22686358]
82. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. Behav Res Ther 2016 Mar;78:30-42. [doi: 10.1016/j.brat.2016.01.001] [Medline: 26848517]
83. Richards D, Timulak L. Satisfaction with therapist-delivered vs self-administered online cognitive behavioural treatments for depression symptoms in college students. Br J Guid Counc 2013 Apr;41(2):193-207. [doi: 10.1080/03069885.2012.726347]
84. Richards D, Timulak L, Hevey D. A comparison of two online cognitive-behavioural interventions for symptoms of depression in a student population: the role of therapist responsiveness. Couns Psychother Res 2013 Sep;13(3):184-193. [doi: 10.1080/14733145.2012.733715]
85. Richards D, Timulak L, Rashleigh C, McLoughlin O, Colla A, Joyce C, et al. Effectiveness of an internet-delivered intervention for generalized anxiety disorder in routine care: a randomised controlled trial in a student population. Internet Interv 2016 Nov;6:80-88 [FREE Full text] [doi: 10.1016/j.invent.2016.10.003] [Medline: 30135817]
86. Santucci LC, McHugh RK, Elkins RM, Schechter B, Ross MS, Landa CE, et al. Pilot implementation of computerized cognitive behavioral therapy in a university health setting. Adm Policy Ment Health 2014 Jul;41(4):514-521. [doi: 10.1007/s10488-013-0488-2] [Medline: 23592231]
87. Seligman ME, Schulman P, Tryon AM. Group prevention of depression and anxiety symptoms. Behav Res Ther 2007 Jun;45(6):1111-1126. [doi: 10.1016/j.brat.2006.09.010] [Medline: 17074301]
88. Tillfors M, Carlbring P, Furmark T, Lewenhaupt S, Spak M, Eriksson A, et al. Treating university students with social phobia and public speaking fears: internet delivered self-help with or without live group exposure sessions. Depress Anxiety 2008;25(8):708-717. [doi: 10.1002/da.20416] [Medline: 18729147]
89. Yang W, Ding Z, Dai T, Peng F, Zhang JX. Attention bias modification training in individuals with depressive symptoms: a randomized controlled trial. J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):101-111 [FREE Full text] [doi: 10.1016/j.jbtep.2014.08.005] [Medline: 25245928]
90. Asbury ET, Casey J, Desai K. Family eJournal: benefits of online guided group journaling for women. J Public Ment Health 2018 Sep 17;17(3):135-141. [doi: 10.1108/JPMH-01-2018-0008]
91. Auyeung L, Mo PK. The efficacy and mechanism of online positive psychological intervention (PPI) on improving well-being among Chinese university students: a pilot study of the best possible self (BPS) intervention. J Happiness Stud 2018 Nov 26:1-6. [doi: 10.1007/s10902-018-0054-4]
92. Bedford LA, Dietch JR, Taylor DJ, Boals A, Zayfert C. Computer-guided problem-solving treatment for depression, PTSD, and insomnia symptoms in student veterans: a pilot randomized controlled trial. Behav Ther 2018 Dec;49(5):756-767. [doi: 10.1016/j.beth.2017.11.010] [Medline: 30146142]
93. Braithwaite SR, Fincham FD. ePREP: computer based prevention of relationship dysfunction, depression and anxiety. J Soc Clin Psychol 2007 May;26(5):609-622. [doi: 10.1521/jscp.2007.26.5.609]
94. Eustis EH, Hayes-Skelton SA, Orsillo SM, Roemer L. Surviving and thriving during stress: a randomized clinical trial comparing a brief web-based therapist-assisted acceptance-based behavioral intervention versus waitlist control for college students. Behav Ther 2018 Dec;49(6):889-903. [doi: 10.1016/j.beth.2018.05.009] [Medline: 30316488]
95. Flett JA, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018 Oct 31;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
96. Freeman E, Barker C, Pistrang N. Outcome of an online mutual support group for college students with psychological problems. Cyberpsychol Behav 2008 Oct;11(5):591-593. [doi: 10.1089/cpb.2007.0133] [Medline: 18817485]
97. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]

98. Gibbel MR. ETD Home. Bowling Green, Ohio: Bowling Green State University; 2010. Evaluating a Spiritually Integrated Intervention for Depressed College Students URL: https://etd.ohiolink.edu/!etd.send_file?accession=bgsu1274920865&disposition=inline
99. Greer C. University Digital Conservancy Home. Minnesota: University of Minnesota; 2015. An Online Mindfulness Intervention to Reduce Stress and Anxiety Among College Students URL: https://conservancy.umn.edu/bitstream/handle/11299/175442/Greer_umn_0130E_16438.pdf?sequence=1&isAllowed=y
100. Hall BJ, Xiong P, Guo X, Sou EK, Chou UI, Shen Z. An evaluation of a low intensity mHealth enhanced mindfulness intervention for Chinese university students: a randomized controlled trial. Psychiatry Res 2018 Dec;270:394-403. [doi: 10.1016/j.psychres.2018.09.060] [Medline: 30300870]
101. Harrer M, Adam SH, Fleischmann RJ, Baumeister H, Auerbach R, Bruffaerts R, et al. Effectiveness of an internet- and app- based intervention for college students with elevated stress: randomized controlled trial. J Med Internet Res 2018 Dec 23;20(4):e136 [FREE Full text] [doi: 10.2196/jmir.9293] [Medline: 29685870]
102. Lee RA, Jung ME. Evaluation of an mHealth app (DeStressify) on university students' mental health: pilot trial. JMIR Ment Health 2018 Jan 23;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
103. McCall HC, Richardson CG, Helgadottir FD, Chen FS. Evaluating a web-based social anxiety intervention among university students: randomized controlled trial. J Med Internet Res 2018;20(3):e91 [FREE Full text] [doi: 10.2196/jmir.8630] [Medline: 29563078]
104. Mogoaşe C, Brăilean A, David D. Can concreteness training alone reduce depressive symptoms? A randomized pilot study using an internet-delivered protocol. Cognit Ther Res 2013 Jan 3;37(4):704-712. [doi: 10.1007/s10608-012-9514-z]
105. Rose RD, Buckey 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]
106. Sagon AL, Danitz SB, Suvak MK, Orsillo SM. The mindful way through the semester: evaluating the feasibility of delivering an acceptance-based behavioral program online. J Contextual Behav Sci 2018 Jul;9:36-44. [doi: 10.1016/j.jcbs.2018.06.004]
107. Saleh D, Camart N, Sbeira F, Romo L. Can we learn to manage stress? A randomized controlled trial carried out on university students. PLoS One 2018;13(9):e0200997 [FREE Full text] [doi: 10.1371/journal.pone.0200997] [Medline: 30183710]
108. Sarniak RH. Carolina Digital Repository. Chapel Hill, NC: The University of North Carolina at Chapel Hill; 2009. The effects of an online intervention designed to cultivate positive emotions on emotional and health outcomes in college students URL: https://cdr.lib.unc.edu/downloads/9593tv255?locale=en
109. Stallman HM, Ohan JL, Chiera B. Reducing distress in university students: a randomised control trial of two online interventions. Aust Psychol 2018 Nov 19;54(2):125-131. [doi: 10.1111/ap.12375]
110. Taitz I. Learning lucid dreaming and its effect on depression in undergraduates. Int J Dream Res 2011;4(2):117-126. [doi: 10.11588/ijodr.2011.2.9123]
111. Villani D, Riva D. Presence and relaxation: a preliminary controlled study. PsychNol J 2008;6(1):7-25 [FREE Full text]
112. Howell AN, Rheingold AA, Uhde TW, Guille C. Web-based CBT for the prevention of anxiety symptoms among medical and health science graduate students. Cogn Behav Ther 2018 Dec 11:1-21 (forthcoming). [doi: 10.1080/16506073.2018.1533575] [Medline: 30526364]
113. Benton SA, Heesacker M, Snowden SJ, Lee G. Therapist-assisted, online (TAO) intervention for anxiety in college students: TAO outperformed treatment as usual. Prof Psychol Res Pr 2016;47(5):363-371. [doi: 10.1037/pro0000097]
114. Carey TA, Haviland J, Tai SJ, Vanags T, Mansell W. MindSurf: a pilot study to assess the usability and acceptability of a smartphone app designed to promote contentment, wellbeing, and goal achievement. BMC Psychiatry 2016 Dec 12;16(1):442 [FREE Full text] [doi: 10.1186/s12888-016-1168-z] [Medline: 27955643]
115. Finlay-Jones A, Kane R, Rees C. Self-compassion online: a pilot study of an internet-based self-compassion cultivation program for psychology trainees. J Clin Psychol 2017 Jul;73(7):797-816. [doi: 10.1002/jclp.22375] [Medline: 27787877]
116. Haas A, Koestner B, Rosenberg J, Moore D, Garlow SJ, Sedway J, et al. An interactive web-based method of outreach to college students at risk for suicide. J Am Coll Health 2008;57(1):15-22. [doi: 10.3200/JACH.57.1.15-22] [Medline: 18682341]
117. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. Arch Psychiatr Nurs 2013 Apr;27(2):84-89. [doi: 10.1016/j.apnu.2012.12.005] [Medline: 23540518]
118. Kaczmarek LD, Kashdan TB, Kleiman EM, Baczkowski B, Enko J, Siebers A, et al. Who self-initiates gratitude interventions in daily life? An examination of intentions, curiosity, depressive symptoms, and life satisfaction. Pers Individ Dif 2013 Oct;55(7):805-810. [doi: 10.1016/j.paid.2013.06.013]
119. Kim EH, Coumar A, Lober WB, Kim Y. Addressing mental health epidemic among university students via web-based, self-screening, and referral system: a preliminary study. IEEE Trans Inf Technol Biomed 2011 Mar;15(2):301-307. [doi: 10.1109/TITB.2011.2107561] [Medline: 21257386]
120. Levin ME, Pistorello J, Hayes SC, Seeley JR, Levin C. Feasibility of an acceptance and commitment therapy adjunctive web-based program for counseling centers. J Couns Psychol 2015 Jul;62(3):529-536. [doi: 10.1037/cou0000083] [Medline: 25961755]

121. Moir F, Fernando AT, Kumar S, Henning M, Moyes SA, Elley CR. Computer assisted learning for the mind (CALM): the mental health of medical students and their use of a self-help website. N Z Med J 2015 Mar 27;128(1411):51-58. [Medline: 25820503]
122. North MM, Schoeneman CM, Mathis JR. Virtual reality therapy: case study of fear of public speaking. Stud Health Technol Inform 2002;85:318-320. [doi: 10.3233/978-1-60750-929-5-318] [Medline: 15458108]
123. Sharry J, Davidson R, McLoughlin O, Doherty G. A service-based evaluation of a therapist-supported online cognitive behavioral therapy program for depression. J Med Internet Res 2013 Jun 27;15(6):e121 [FREE Full text] [doi: 10.2196/jmir.2248] [Medline: 23807565]
124. Spadaro KC, Hunker DF. Exploring the effects of an online asynchronous mindfulness meditation intervention with nursing students on stress, mood, and cognition: a descriptive study. Nurse Educ Today 2016 Apr;39:163-169. [doi: 10.1016/j.nedt.2016.02.006] [Medline: 27006051]
125. Trockel M, Manber R, Chang V, Thurston A, Taylor CB. An e-mail delivered CBT for sleep-health program for college students: effects on sleep quality and depression symptoms. J Clin Sleep Med 2011 Jun 15;7(3):276-281 [FREE Full text] [doi: 10.5664/JCSM.1072] [Medline: 21677898]
126. Williams A, Larocca R, Chang T, Trinh NH, Fava M, Kvedar J, et al. Web-based depression screening and psychiatric consultation for college students: a feasibility and acceptability study. Int J Telemed Appl 2014;2014:580786 [FREE Full text] [doi: 10.1155/2014/580786] [Medline: 24799895]
127. Wilson FR, Omeltschenko L, Yager GG. Coping with test stress: microcomputer software for treatment of test anxiety. J Behav Ther Exp Psychiatry 1991 Jun;22(2):131-139. [doi: 10.1016/0005-7916(91)90008-S] [Medline: 1684587]
128. Palacios JE, Richards D, Palmer R, Coudray C, Hofmann SG, Palmieri PA, et al. Supported internet-delivered cognitive behavioral therapy programs for depression, anxiety, and stress in university students: open, non-randomised trial of acceptability, effectiveness, and satisfaction. JMIR Ment Health 2018 Dec 14;5(4):e11467 [FREE Full text] [doi: 10.2196/11467] [Medline: 30552076]
129. Fleischmann R, Harrer M, Zarski AC, Baumeister H, Lehr D, Ebert DD. Patients' experiences in a guided internet- and app-based stress intervention for college students: a qualitative study. Internet Interv 2018 Jun;12:130-140 [FREE Full text] [doi: 10.1016/j.invent.2017.12.001] [Medline: 30135777]
130. García-Magariño I, Bedia MG, Palacios-Navarro G. FAMAP: A Framework for Developing m-Health Apps. In: Proceedings of the Trends and Advances in Information Systems and Technologies. 2018 Presented at: WorldCIST'18; March 27-29, 2018; Naples, Italy p. 850-859. [doi: 10.1007/978-3-319-77703-0_83]
131. Henriques G, Keffer S, Abrahamson C, Horst SJ. Exploring the effectiveness of a computer-based heart rate variability biofeedback program in reducing anxiety in college students. Appl Psychophysiol Biofeedback 2011 Jun;36(2):101-112. [doi: 10.1007/s10484-011-9151-4] [Medline: 21533678]
132. Brooke J. SUS-a quick and dirty usability scale. J Usability Stud 1996;189(194):4-7 [FREE Full text]
133. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-X] [Medline: 10259963]
134. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Dec;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
135. Ebert DD, Zarski AC, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
136. Fortney JC, Curran GM, Hunt JB, Cheney AM, Lu L, Valenstein M, et al. Prevalence of probable mental disorders and help-seeking behaviors among veteran and non-veteran community college students. Gen Hosp Psychiatry 2016;38:99-104 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.09.007] [Medline: 26598288]
137. Fortney JC, Curran GM, Hunt JB, Lu L, Eisenberg D, Valenstein M. Mental health treatment seeking among veteran and civilian community college students. Psychiatr Serv 2017 Aug 1;68(8):851-855. [doi: 10.1176/appi.ps.201600240] [Medline: 28412894]
138. Sontag-Padilla L, Woodbridge MW, Mendelsohn J, D'Amico EJ, Osilla KC, Jaycox LH, et al. Factors affecting mental health service utilization among California public college and university students. Psychiatr Serv 2016 Dec 1;67(8):890-897. [doi: 10.1176/appi.ps.201500307] [Medline: 27032662]
139. Hanel PH, Vione KC. Do student samples provide an accurate estimate of the general public? PLoS One 2016;11(12):e0168354 [FREE Full text] [doi: 10.1371/journal.pone.0168354] [Medline: 28002494]
140. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
141. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010 Nov 23;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]

142. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
143. Stiles-Shields C, Montague E, Lattie EG, Schueller SM, Kwasny MJ, Mohr DC. Exploring user learnability and learning performance in an app for depression: usability study. JMIR Hum Factors 2017 Aug 11;4(3):e18 [FREE Full text] [doi: 10.2196/humanfactors.7951] [Medline: 28801301]
144. Vilardaga R, Rizo J, Kientz JA, McDonell MG, Ries RK, Sobel K. User experience evaluation of a smoking cessation app in people with serious mental illness. Nicotine Tob Res 2016 May;18(5):1032-1038 [FREE Full text] [doi: 10.1093/ntr/ntv256] [Medline: 26581430]
145. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
146. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
147. Stiles-Shields C, Holmbeck GN. JPP student journal club commentary: smartphone-delivered interventions for pediatric populations: improving methodologies to address concerns of feasibility and efficacy. J Pediatr Psychol 2018 Jun 1;43(5):485-487 [FREE Full text] [doi: 10.1093/jpepsy/jsy018] [Medline: 29546284]
148. Lattie EG, Ho J, Sargent E, Tomasino KN, Smith JD, Brown CH, et al. Teens engaged in collaborative health: the feasibility and acceptability of an online skill-building intervention for adolescents at risk for depression. Internet Interv 2017 Jun;8:15-26 [FREE Full text] [doi: 10.1016/j.invent.2017.02.003] [Medline: 28584734]
149. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Dec 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
150. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. eMental healthcare technologies for anxiety and depression in childhood and adolescence: systematic review of studies reporting implementation outcomes. JMIR Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]
151. 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]
152. 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]
153. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]


## Abbreviations

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RCT:** randomized controlled trial


*Edited by G Eysenbach; submitted 20.11.18; peer-reviewed by E Montague, K Mishina, N Gehring, EB Davies, A Newton; comments to author 17.04.19; revised version received 29.05.19; accepted 14.06.19; published 22.07.19.*

*Please cite as:*
*Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK*
*Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review*
*J Med Internet Res 2019;21(7):e12869*
*URL: http://www.jmir.org/2019/7/e12869/*
*doi:10.2196/12869*
*PMID:31333198*

Original Paper

# Identification of Patients in Need of Advanced Care for Depression Using Data Extracted From a Statewide Health Information Exchange: A Machine Learning Approach

Suranga N Kasthurirathne[1,2], PhD; Paul G Biondich[1,3], MD, MS; Shaun J Grannis[1,3], MD, MS; Saptarshi Purkayastha[4], PhD; Joshua R Vest[1,2], PhD, MPH; Josette F Jones[4], PhD

[1]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States
[2]Indiana University Fairbanks School of Public Health, Indianapolis, IN, United States
[3]Indiana University School of Medicine, Indianapolis, IN, United States
[4]Indiana University School of Informatics and Computing, Indianapolis, IN, United States

**Corresponding Author:**
Suranga N Kasthurirathne, PhD
Center for Biomedical Informatics
Regenstrief Institute
1101 W 10th St
Indianapolis, IN, 46202
United States
Phone: 1 3173323480
Email: snkasthu@iupui.edu

## Abstract

**Background:** As the most commonly occurring form of mental illness worldwide, depression poses significant health and economic burdens to both the individual and community. Different types of depression pose different levels of risk. Individuals who suffer from mild forms of depression may recover without any assistance or be effectively managed by primary care or family practitioners. However, other forms of depression are far more severe and require advanced care by certified mental health providers. However, identifying cases of depression that require advanced care may be challenging to primary care providers and health care team members whose skill sets run broad rather than deep.

**Objective:** This study aimed to leverage a comprehensive range of patient-level diagnostic, behavioral, and demographic data, as well as past visit history data from a statewide health information exchange to build decision models capable of predicting the need of advanced care for depression across patients presenting at Eskenazi Health, the public safety net health system for Marion County, Indianapolis, Indiana.

**Methods:** Patient-level diagnostic, behavioral, demographic, and past visit history data extracted from structured datasets were merged with outcome variables extracted from unstructured free-text datasets and were used to train random forest decision models that predicted the need of advanced care for depression across (1) the overall patient population and (2) various subsets of patients at higher risk for depression-related adverse events; patients with a past diagnosis of depression; patients with a Charlson comorbidity index of ≥1; patients with a Charlson comorbidity index of ≥2; and all unique patients identified across the 3 above-mentioned high-risk groups.

**Results:** The overall patient population consisted of 84,317 adult (aged ≥18 years) patients. A total of 6992 (8.29%) of these patients were in need of advanced care for depression. Decision models for high-risk patient groups yielded area under the curve (AUC) scores between 86.31% and 94.43%. The decision model for the overall patient population yielded a comparatively lower AUC score of 78.87%. The variance of optimal sensitivity and specificity for all decision models, as identified using Youden J Index, is as follows: sensitivity=68.79% to 83.91% and specificity=76.03% to 92.18%.

**Conclusions:** This study demonstrates the ability to automate screening for patients in need of advanced care for depression across (1) an overall patient population or (2) various high-risk patient groups using structured datasets covering acute and chronic conditions, patient demographics, behaviors, and past visit history. Furthermore, these results show considerable potential to enable preventative care and can be easily integrated into existing clinical workflows to improve access to wraparound health care services.

*(J Med Internet Res 2019;21(7):e13809)* doi:10.2196/13809

**KEYWORDS**

depression; supervised machine learning; delivery of health care

## Introduction

### Background

Depression is the most commonly occurring mental illness worldwide [1]. It negatively affects how up to 350 million persons worldwide think, feel, and interact [2]. Depression poses significant health and economic burdens to both the individual and community [3]. Previous studies have presented a strong comorbidity between mental health and medical conditions [4]. Depression is highly prevalent among patients suffering from various chronic conditions [5,6]. Such patients may suffer up to a 10-to-25-year reduction in life expectancy [7,8]. Depression is also a leading cause of disability for Americans aged between 15 and 44 years [9]. The incremental economic burden of depression covering medical, pharmaceutical, workplace, and suicide-related costs in the United States was evaluated at US $210.5 billion in 2010, a 21.5% increase from 2005 [10].

Different types of depression pose different levels of risk. Individuals who suffer from mild forms of depression may recover without any assistance. Other less severe cases can be effectively managed by primary care or family practitioners [11-13]. However, other forms of depression are far more severe and require advanced care above and beyond that provided by primary care or family practitioners [14,15]. Identifying cases of depression that require advanced care may be challenging to primary care providers and health care team members whose skill sets run broad rather than deep. Training health care teams to successfully identify patients with severe depression would resolve the problem but is unfeasible given cost, effort, and time considerations [16,17]. Social stigma and ignorance of health issues also encourage depression sufferers to downplay their condition, further increasing difficulty in detection and assessment [18].

Many health care systems leverage screening tools such as the Beck Depression Scale [19], the Patient Health Questionnaire-9 (PHQ-9) [20], PHQ-15 [21], the Cornell Scale for Depression in Dementia [22], and the Hamilton Rating Scale for Depression [23] to evaluate depression severity. However, such tools are not optimal as they (1) tie up significant resources [24], (2) rely heavily on potentially inaccurate patient-reported outcomes for decision making [25], and (3) utilize only a small subset of clinical and behavioral data for decision making. In addition, traditional depression screening approaches may increase risk of overdiagnosis and overtreatment of depression across community and primary care settings [26-28] without contributing to better mental health [29]. Recent studies have questioned the benefits of routine screening [30,31] as well as the US Preventive Services Task Force recommendations to screen adults for depression in primary care settings where staff-assisted depression management programs are available [29].

Given such limitations, it is more appropriate to develop machine learning–based screening approaches capable of leveraging more comprehensive patient datasets representing a patient's overall health status to identify individuals who cannot be treated at primary care alone and would suffer from worsening health conditions unless they are provided with specialized, high-intensity treatment for depression [14,15]. Machine learning enables us to learn from multiple primary and secondary care datasets that might be missed by a clinician because of cognitive burden, and therefore, are a suitable solution to this challenge.

### Objectives

For purposes of this research, we have defined individuals whose quality of life and health status will degrade if they do not receive specialized treatment above and beyond primary care as patients in need of advanced care for depression. Operationally, such patients would be identified by evaluating clinical data to detect patients who had received referrals to a certified mental health provider for specialized treatment for depression, indicating that their illnesses cannot be treated at primary care alone. In this study, we leveraged data obtained from varied structured and unstructured datasets to build decision models capable of identifying patients in need of advanced care for depression.

## Methods

### Patient Population

We identified a population of 84,317 adult patients (≥18 years of age) with at least 1 primary care visit between the years 2011 and 2016 at Eskenazi Health, a leading health care provider in Indianapolis, Indiana.

### Patient Subset Selection

We sought to predict the need for advanced care for depression across (1) the overall patient population and (2) different groups of high-risk patient populations. We selected 3 high-risk patient groups: group A: patients with a past diagnosis of depression, group B: patients with a Charlson Comorbidity Index [32] of ≥1, and group C: patients with a Charlson Comorbidity Index of ≥2. Patients with a past diagnosis of depression were flagged as a high-risk group as their illness may re-emerge or worsen based on other health conditions. Patients with Charlson indexes ≥1 and ≥2 were selected because of the high prevalence of depression among patients suffering from one or many chronic illnesses [33] and its ability to worsen health outcomes of patients. Thresholds of ≥1 and ≥2 were selected because they captured patient populations that were adequately large for machine learning processes, as well as the cost/effort of potential implementation. We also identified a fourth group (Group D) that comprised all unique patients identified in groups A to C.

We trained models for different populations to capture as many of the overall number of patients in need of advanced care for depression and to identify which patient groups were most suitable for use in screening for need of advanced care. Furthermore, focusing on a smaller population of high-risk patients may be easier to operationalize and cost-efficient to implement across chronic care clinics. Groups A to D were identified by analyzing diagnostic data on each of the 84,317 unique patients (master patient list) for past diagnosis of depression and to calculate Charlson Comorbidity Index for each patient.

### Data Preparation

In a previous effort, we developed a depression taxonomy [34] using knowledge-based terminology extraction of the Unified Medical Language System (UMLS) Metathesaurus [35]. The taxonomy was developed by performing a literature search on Ovid Medline to identify publications that discuss depression and its treatment and then using Metamap [36], a Natural Language Processing–based tool to map these abstracts against the UMLS Metathesaurus, a large, multipurpose, multilingual thesaurus that contains millions of biomedical and health-related concepts, synonymous names, and their relationships across 199 medical dictionaries [37]. The most frequently occurring UMLS concepts were compiled into a terminology using the Web Ontology Language, a semantic Web language that is widely used to represent ontologies. These features presented a wide variety of diagnostic, demographic, and behavioral features that impacted the onset and severity of depression [34].

We obtained longitudinal health records on each patient from the Indiana Network for Patient Care (INPC), a statewide health information exchange [38,39]. Thus, our dataset included records on each patient, including data that may have been captured at any hospital system that participated in the INPC. The dataset included a wide array of patient data, including patient demographic, diagnostic, behavioral, and visit data reported in both structured and unstructured form. All diagnostic data were obtained in the form of structured International Classification of Diseases, ninth revision (ICD-9) and ICD-10 codes. We assessed extracted data against the depression terminology and used relationships presented within the UMLS Metathesaurus to identify ICD codes for inclusion as features. We tabulated vectors of features for each patient group under study. We predict current risk levels based on past patient data. We did not assess the impact of temporality because of our dataset representing a (1) relatively short time period and (2) an older population with high chronic conditions that do not change significantly over time. In the event that the patient under study had received a referral for depression treatment, the data vector only comprised medical data recorded up to 24 hours before the aforesaid referral order. If no past referrals for depression treatment were present, then the vector comprised all available data on the patient. A master data vector encompassing all 84,317 patients was also created using the same approach.

### Preparation of Gold Standard

We applied regular expressions to physician referrals to certified mental health providers to identify referrals where the physician was recommending specialized treatment for depression. We determined that our use of regex patterns was 100% accurate via manual review.

### Decision Model Building

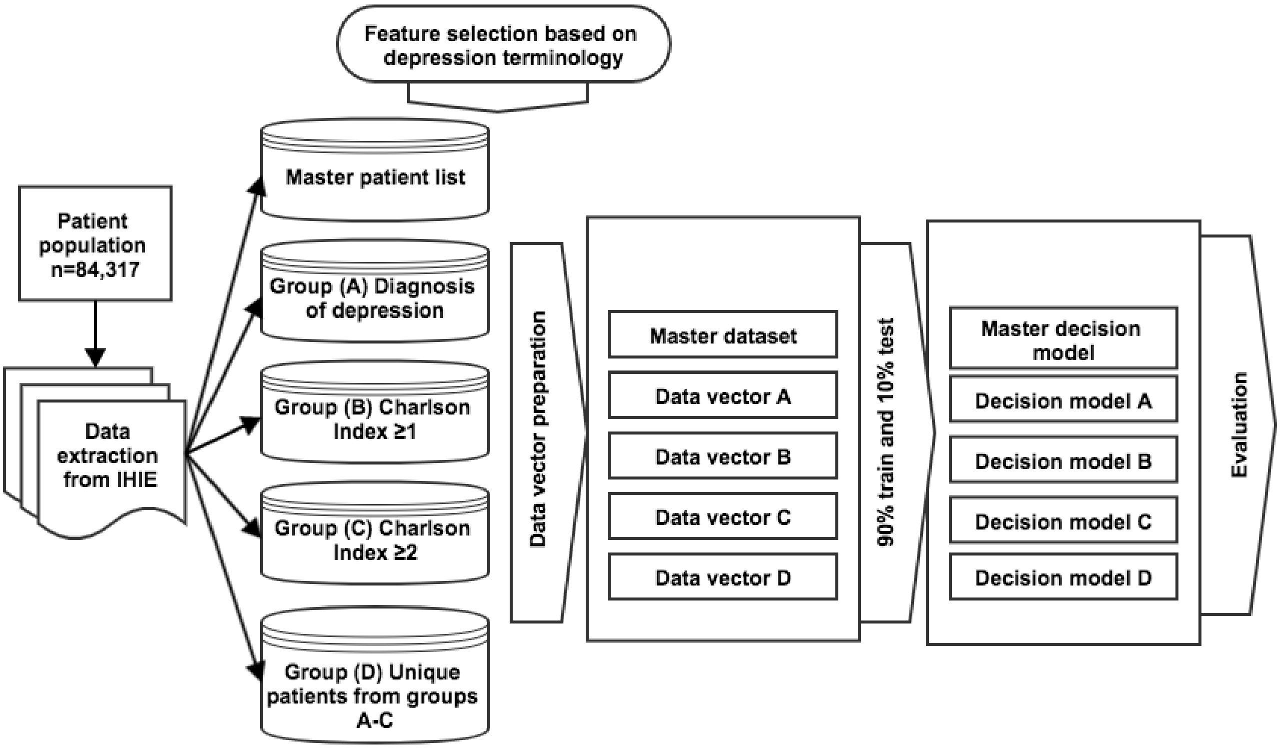
We split each of the 5 data vectors (4 patient subgroups and 1 master data vector) into random groups of 90% training data and 10% test data. Each training dataset was used to train a decision model using the random forest classification algorithm [40]. The random forest algorithm was selected because of its track record of successful use in decision modeling for health care challenges [41,42] and its ability to develop interpretable machine learning predictions [43]. We used Python programming language (version 2.7.6) for all data preprocessing tasks and the Python scikit-learn package for decision model development and testing [44].

### Analysis

Each decision model was evaluated using the 10% holdout test set. Results produced by each decision model were evaluated using area under the curve (AUC) values, which measure classifier accuracy. Youden J Index [45] was used to identify optimal sensitivity and specificity for each decision model.

A flowchart representing our workflow from patient group selection to decision model evaluation can be seen in Figure 1.

**Figure 1.** A flowchart representing our workflow from patient group selection to decision model evaluation. IHIE: Indiana Health Information Exchange.




## Results

### Evaluation of Patient Groups

We identified a total of 12,432 patients with a diagnosis of depression (group A), 32,249 patients with a Charlson Index of 1 or greater (group B), and 7415 patients with a Charlson Index of 2 or more (group C). Overall, these 3 groups identified a total of 37,560 unique patients (group D).

The master patient list as well as each of the 4 high-risk patient groups represented an adult, urban population: predominantly female and with high disease burdens (Table 1). The populations identified by their Charlson Indexes were older (mean age >50 years) than the population identified with depression (46.31 mean age). In addition, populations identified based on Charlson Indexes were predominantly African American. In contrast, the population with a past diagnosis of depression was predominantly composed of non-Hispanic whites. As anticipated, the prevalence of depression across a patient population with a Charlson Index of 1 or greater (30.18%) and a patient population with a Charlson Index of 2 or greater (37.25%) was greater than across the master patient list (19%).

Figure 2 presents a Venn diagram presenting overlap across the high-risk patient groups identified for the study.

A total of 6992 (8.29%) of the 84,317 patients in the master patient list were in need of advanced care for depression. Group A captured 3683 (52.68%) of these patients, group B captured 4016 (57.43%), and group C captured 1026 (14.67%). Overall, all 3 patient groups were able to identify 5612 (80.26%) of all patients in need of advanced care for depression.

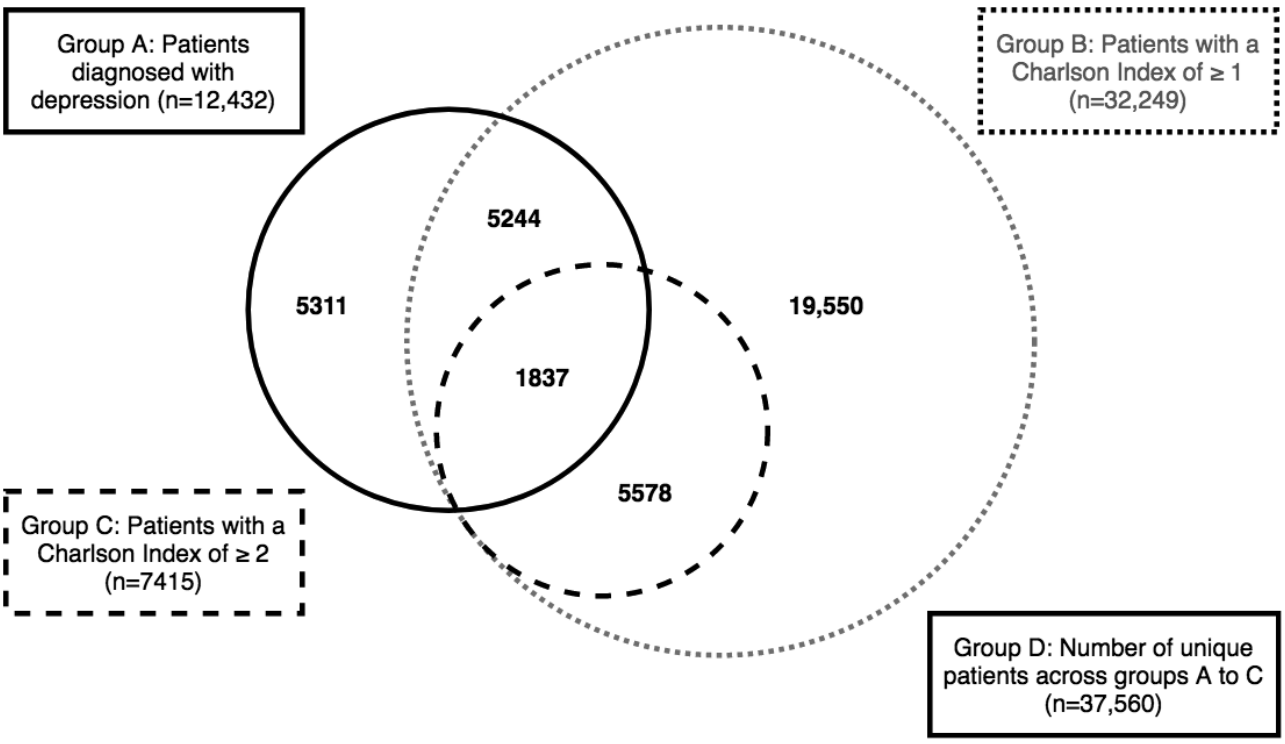
**Table 1.** Characteristics of the master patient list/groups of high-risk patients used for decision model building.

| Characteristic of interest | Master patient set: all patients (N=84,317) | Group A: patients with a past diagnosis of depression | Group B: patients with a Charlson Index of ≥1 | Group C: patients with a Charlson Index of ≥2 | Group D: all unique patients in groups A-C |
|---|---|---|---|---|---|
| Patient group size, n (%) | __[a] | 12,432 (14.74) | 32,249 (38.25) | 7415 (8.8) | 37,560 (44.5) |
| Need of advanced care for depression, n (%) | 6992 (8.29) | 3683 (30.04) | 4016 (12.94) | 1026 (21.6) | 5612 (80.26) |
| **Demographics** | | | | | |
| Age (years), mean (SD) | 43.88 (15.60) | 46.31 (14.74) | 51.94 (14.55) | 59.50 (12.33) | 50.31 (14.93) |
| Male gender (%) | 35.09 | 30.22 | 39.8 | 43.98 | 42.03 |
| **Race/ethnicity (%)** | | | | | |
| White (non-Hispanic) | 25.21 | 44.62 | 33.38 | 37.02 | 35.31 |
| African American (non-Hispanic) | 37.23 | 32.01 | 42.78 | 47.26 | 40.12 |
| Hispanic or Latino | 19.47 | 11.12 | 10.60 | 4.94 | 7.38 |
| **Diagnoses** | | | | | |
| Depression (%) | 19.07 | 100 | 30.18 | 37.25 | 37.51 |
| Charlson Index score, mean (SD) | 0.77 (1.21) | 0.22 (0.75) | 1.89 (1.27) | 3.85 (1.14) | 1.62 (1.35) |
| **Hospitalizations, mean (SD)** | | | | | |
| ED[b] visits during current month | 0.21 (1.03) | 0.33 (1.48) | 0.26 (1.15) | 0.31 (1.14) | 0.27 (1.17) |
| ED visits before previous months | 3.73 (14.40) | 4.69 (18.73) | 8.63 (24.2) | 10.71 (31.36) | 8.03 (23.67) |

[a]Not applicable.

[b]ED: emergency department.

**Figure 2.** Overlap between the patient groups identified for the study.



### Feature Selection Using the Depression Terminology

Comparison of patient data against the depression terminology resulted in the identification of 1150 unique concepts for inclusion in each decision model. A description of features included in each of the decision models is presented in Multimedia Appendix 1.

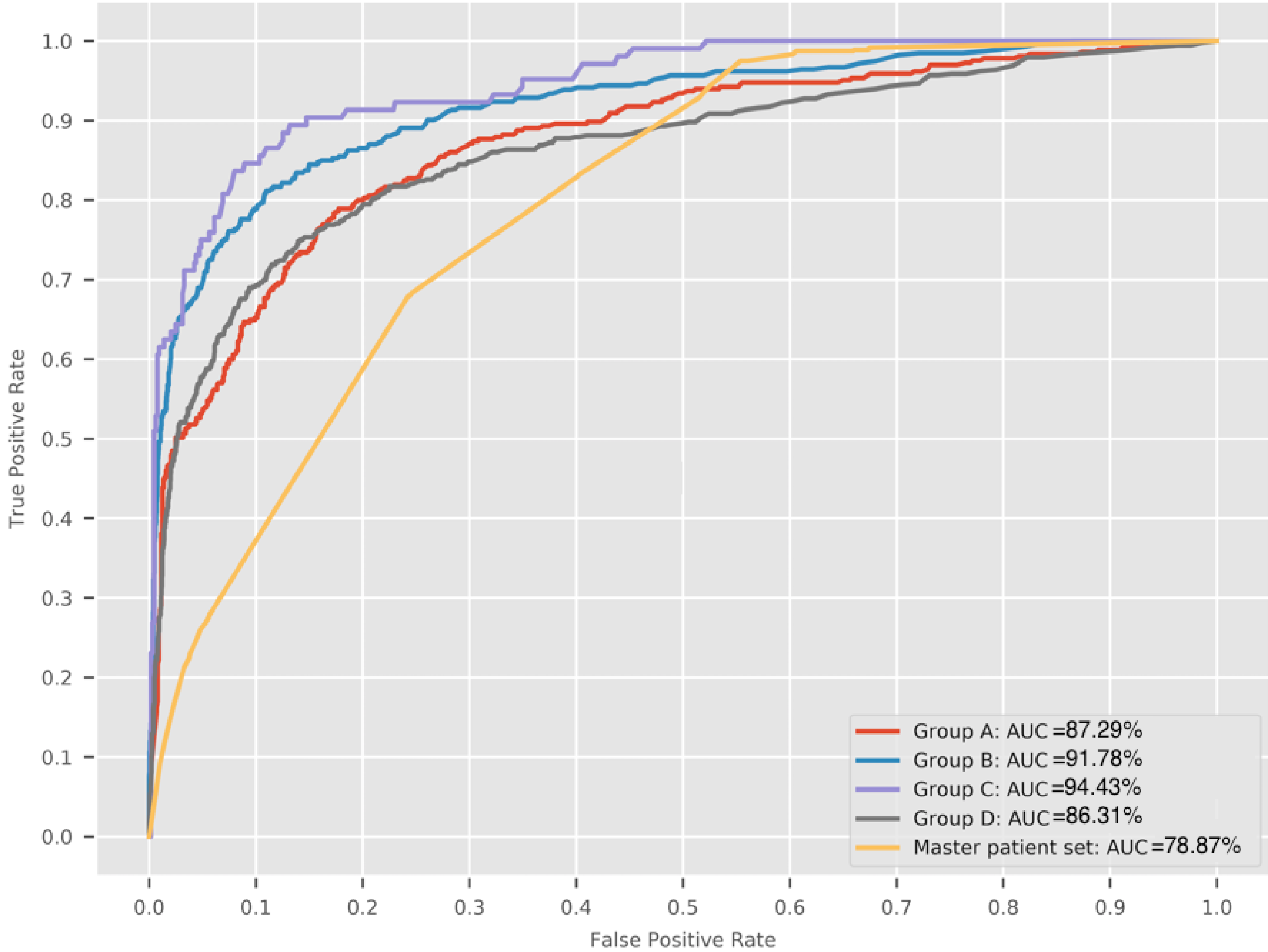
### Decision Model Performance

The decision model predicting need of advanced care across the master population reported a moderate AUC score of 78.87% (optimal sensitivity=68.79%, optimal specificity=76.30%). However, decision models to predict need of advanced care across patients' groups A to D performed significantly better. Group A (patients with a past diagnosis of depression) reported an AUC score of 87.29% (optimal sensitivity=77.84%, optimal specificity=82.66%). Group B (patients with a Charlson Index of ≥1) reported an AUC score of 91.78% (optimal sensitivity=81.05%, optimal specificity=89.21%). Group C (patients with a Charlson Index of ≥ 2) reported an AUC score of 94.43% (optimal sensitivity=83.91%, optimal specificity=92.18%), whereas Group D (list of unique patients from groups A-C) reported an AUC score of 86.31% (Figure 3; optimal sensitivity=75.31%, optimal specificity=76.03%). Precision-recall curves for each decision model are presented in Multimedia Appendix 2.

The top 20 features for each decision model can be seen in Multimedia Appendix 3. Multimedia Appendix 4 presents the co-occurrence of these top 20 features across each decision model under study. In assessing the top ranked features selected for each decision model, we found significant overlap among the top features for each of the high-risk patient populations. Furthermore, essential (primary) hypertension, depression, gender, and number of outpatient visits appears in the top 20 feature lists for every patient population under test.

To demonstrate that the models did not suffer from overtraining, we added an additional evaluation step where we compared model performance across smaller feature subset sizes. We ranked all features for each decision model using information gain aka. Kullback-Leibler divergence [46]. For each patient subgroup, we used the ranked feature lists to build multiple decision models starting with a decision model trained using only the 5 top ranking features, iteratively adding on the next most important feature, retraining the model and evaluating performance using F1 core. We continued this process until we had trained n-5 models using all n features in the feature set. As an example, for patient group A, we began by building a decision model consisting of 5 patient-centric features and assessing its performance using F1 score. Afterwards, we added in the 6th most important feature and retrained a decision model consisting of these 6 features. We continued building models and evaluating F1 scores until we had included all features from each dataset. The results of this exercise (Multimedia Appendix 5) demonstrated that model performance plateaued after the top 10 to 20 features and that inclusion of further features did not improve model performance. This demonstrates that the models were not overfit and that they reached optimal performance after a relatively small number of features.

**Figure 3.** Receiver operating characteristic curves produced by decision models predicting need of advanced care across each patient group under study. AUC: area under the curve.




## Discussion

### Principal Findings

The decision model to predict need of advanced care for depression across the overall patient population achieved an AUC score of 78.87%. In comparison, decision models that predicted need of advanced care across 4 high-risk patient groups performed better, with AUC scores ranging from 86.31 to 94.43%. In addition, optimal sensitivity and specificity for each decision model was significantly high and demonstrated the models' potential for practical implementation.

We attribute the comparatively lower performance of the decision model developed using the overall population to the unbalanced nature of the gold standard [47] caused by the relatively low prevalence (8.29%) of patients in need of advanced care and the sparsity of data available for some of the patients in the overall patient population. The high performance of the decision models built using high-risk patient groups could be attributed to the higher prevalence of patients in need of advanced care. Although various publications have presented approaches to address data imbalance [48,49], we did not pursue such as approach as we wished to focus on demonstrating methods that could be replicated across other datasets that may or may not be imbalanced.

In assessing prediction performance, group C (patients with a Charlson Index ≥2) yielded the highest AUC score (97.43%). Groups A (patients with a diagnosis of depression) and B (patients with a Charlson Index ≥1) reported lesser AUC scores. Group C captured the least number of patients in need of advanced care in comparison with groups A and B. However, it is noteworthy that none of the decision models developed using high-risk populations could capture all patients in need of advanced care. Overall, all 3 models could capture only 80.26% of all patients in need of advanced care. The remainder (19.74%) of the patients in need of advanced care did not qualify for any of the three high-risk patient populations. We hypothesize that a share of the missing 19.74% patients would have fallen into 1 of the 3 high-risk patient groups had more comprehensive data been available, and thus, been eligible for detection.

We present a novel application of machine learning to address a question of significant clinical relevance. We demonstrated the ability to predict the need of advanced care for depression across various patient populations with considerable predictive performance. These efforts can easily be integrated into existing hospital workflows [42]. As wraparound services are not delivered by primary care providers [50], the ability to identify and refer patients in need of such services is extremely important [51]. Our efforts yield a highly accurate, automated approach for identifying patients in need of wraparound services for mental health, which is of growing importance to health care organizations and incentivized by changing reimbursement policies. By predicting the need for advanced care across various high-risk populations, we offer potential implementers the option of selecting the best screening approach that meets their needs. Our approach is also well suited to leverage increasing health information technology adoption and interoperability of health care datasets for community-wide health transformations [52,53]. In the field of population health informatics, it enables organizations to leverage widespread acceptance and use of machine learning for cross organizational collaboration and management of various datasets [53] while giving implementers the freedom to select methods best suited for each hospital system. Furthermore, such applications of predictive modeling could support organization-level population health initiatives as risk stratification is fundamental to identifying those patients who are most in need of services to improve health and well-being. In addition, implementing such solutions at primary care ensure that facilitating the entry of all patients into the health care system is more efficient than stand-alone implementations at each chronic care clinic. Thus, our approach presents the ability to effectively identify need of advanced care for depression without risk of overdiagnosis and overtreatment and without the use of manual screening mechanisms.

There is limited knowledge on the best approach to integrate machine learning approaches into existing clinical workflows. As highlighted above, primary care facilities are the point of entry by which a majority of patients suffering from depression seek care [54,55]. However, application of machine learning solutions to screen every primary care visit may be cost-intensive and inefficient for certain clinical practices. Thus, alternate models to evaluate a subset of high-risk patients in need of advanced care for depression would be useful. The 2 potential high-risk patient populations are (1) patients already diagnosed with depression and (2) patients with chronic conditions, who are thus, are at higher risk of suffering from depression [56]. Models built using these subsets may be more practical and result in better machine learning performance than models built using all primary care patients because of variability of underlying data and higher prevalence of outcome of interest, which enables better model training.

We identified several limitations in our study approach. We adopted a binary (present/absent) flag for each feature used to train decision models. We hypothesize that switching to tabulated counts for each feature will increase the granularity of the feature vector, thereby increasing model performance. The patient group used in our study was obtained from the Eskenazi Health system, a safety-net population with significant health burdens. Thus, our models may not generalize to other commercially insured or broader populations. Our diagnostic data were limited to ICD codes. Integrating medications, laboratory, and clinical procedure data may further improve decision model performance. Furthermore, studies present that social determinants of health such as low-educational attainment, poverty, unemployment, and social isolation may have a significant impact on depression and the need for treatment [57,58]. We propose to expand our models using social determinants of health to assess their impact on decision model performance.

We acknowledge that incompleteness of EMR data [59] may impact model performance. Use of claims data may have enabled us to identify a greater number of patients in need of specified treatment into each of our patient subgroups [60]. Furthermore, our outcome of interest are patients in need of advanced care, as identified by primary care providers. Thus, we were unable to account for patients who received advanced care for depression without a past referral. Such patients could have been identified from claims data and used to augment our gold standard.

We selected the random forest classification algorithm for decision-model building based on the need to develop high performance models that were easily interpretable to our clinical audience [42,43]. Other, more advanced decision-modeling approaches such as neural networks [61] have shown potential to improve machine learning performance across various health care challenges. However, neural networks are more complex, cost-intensive, and difficult to interpret [62], making it harder to gain provider acceptance of such models. In addition, it is unclear if they can contribute to our study given the significant performance measures already achieved using random forest models. We recommend that neural networks be considered in a scenario where the sequence or temporality of clinical events is being evaluated, or where the performance of random forest models is unsatisfactory.

In conclusion, these results present considerable potential to enable preventative care and can be potentially integrated into existing clinical workflows to improve access to wraparound health care services.

## Conclusions

Our efforts demonstrate the ability to identify patients in need of advanced care for depression across (1) an overall patient population and (2) various groups of high-risk patients using a wide range of acute and chronic conditions, patient demographics, behaviors, and past visit history. Although all models yielded significant performance accuracy, models focused on high-risk patient populations yielded comparatively better results. Furthermore, our methods present a replicable approach for implementers to adopt based on their own needs and priorities. However, decision model performance may differ based on the availability of patient data at each health care system. These results show considerable potential to enable preventative care and can be easily integrated into existing clinical workflows to improve access to wraparound health care services.

## Acknowledgments

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

## Authors' Contributions

The primary author proposed and completed the study as part of his doctoral dissertation. The coauthors, who are all members of his dissertation committee, contributed significantly to the acquisition, analysis, and interpretation of data, rigor of the methodology and analysis, as well as contributing to write the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Types of patient data used in decision-model building.

[DOCX File, 13KB - jmir_v21i7e13809_app1.docx ]

## Multimedia Appendix 2

Precision-recall curve for each decision model under study.

[DOCX File, 1MB - jmir_v21i7e13809_app2.docx ]

## Multimedia Appendix 3

List of 20 top features (ranked in order of best to worst) for each decision model, together with their least absolute shrinkage and selection operator scores.

[DOCX File, 16KB - jmir_v21i7e13809_app3.docx ]

## Multimedia Appendix 4

Co-occurrence of top 20 features across each of the patient populations under test (1=most important, 20=least important).

[DOCX File, 19KB - jmir_v21i7e13809_app4.docx ]

## Multimedia Appendix 5

F1 scores for each decision model trained using iteratively increasing feature subset sizes. For clarity, our plot includes only models trained using the top 50 features for each patient subgroup.

[PNG File, 102KB - jmir_v21i7e13809_app5.png ]

## References


1. World Health Organization. 2012. Depression: A Global Crisis. World Mental Health Day, October 10 2012 URL: https://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf
2. American Psychiatric Association. 2016. What is Depression? URL: https://www.psychiatry.org/patients-families/depression/what-is-depression [accessed 2018-01-01]
3. Lépine JP, Briley M. The increasing burden of depression. Neuropsychiatr Dis Treat 2011;7(Suppl 1):3-7 [FREE Full text] [doi: 10.2147/NDT.S19617] [Medline: 21750622]
4. Goodell S, Druss BG, Walker ER. Mental disorders and medical comorbidity. Synth Proj Res Synth Rep 2011 Feb(21):1-26 [FREE Full text] [Medline: 21675009]
5. Yohannes A, Willgoss T, Baldwin R, Connolly M. Depression and anxiety in chronic heart failure and chronic obstructive pulmonary disease: prevalence, relevance, clinical implications and management principles. Int J Geriatr Psychiatry 2010 Dec;25(12):1209-1221. [doi: 10.1002/gps.2463] [Medline: 20033905]
6. Nouwen A, Winkley K, Twisk J, Lloyd C, Peyrot M, Ismail K, European Depression in Diabetes (EDID) Research Consortium. Type 2 diabetes mellitus as a risk factor for the onset of depression: a systematic review and meta-analysis. Diabetologia 2010 Dec;53(12):2480-2486 [FREE Full text] [doi: 10.1007/s00125-010-1874-x] [Medline: 20711716]
7. World Health Organization. 2016. Premature Death Among People With Severe Mental Disorders URL: http://www.who.int/mental_health/management/info_sheet.pdf [accessed 2018-03-01]
8. Hawton K, Casañas IC, Haw C, Saunders K. Risk factors for suicide in individuals with depression: a systematic review. J Affect Disord 2013 May;147(1-3):17-28. [doi: 10.1016/j.jad.2013.01.004] [Medline: 23411024]
9. Mathers C, Fat DM, Boerma JT. World Health Organization. 2008. The Global Burden of Disease: 2004 Update URL: https://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/

10. Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry 2015 Feb;76(2):155-162 [FREE Full text] [doi: 10.4088/JCP.14m09298] [Medline: 25742202]
11. Wagner HR, Burns B, Broadhead W, Yarnall K, Sigmon A, Gaynes B. Minor depression in family practice: functional morbidity, co-morbidity, service utilization and outcomes. Psychol Med 2000 Nov;30(6):1377-1390. [doi: 10.1017/S0033291799002998] [Medline: 11097078]
12. Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unützer J, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. J Am Med Assoc 2000 Jan 12;283(2):212-220. [doi: 10.1001/jama.283.2.212] [Medline: 10634337]
13. Williams Jr JW, Barrett J, Oxman T, Frank E, Katon W, Sullivan M, et al. Treatment of dysthymia and minor depression in primary care: a randomized controlled trial in older adults. J Am Med Assoc 2000 Sep 27;284(12):1519-1526. [doi: 10.1001/jama.284.12.1519] [Medline: 11000645]
14. Saloheimo HP, Markowitz J, Saloheimo TH, Laitinen JJ, Sundell J, Huttunen MO, et al. Psychotherapy effectiveness for major depression: a randomized trial in a Finnish community. BMC Psychiatry 2016 May 6;16:131 [FREE Full text] [doi: 10.1186/s12888-016-0838-1] [Medline: 27153942]
15. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008 Dec;76(6):909-922. [doi: 10.1037/a0013075] [Medline: 19045960]
16. Sharp LK, Lipsky MS. Screening for depression across the lifespan: a review of measures for use in primary care settings. Am Fam Physician 2002 Sep 15;66(6):1001-1008 [FREE Full text] [Medline: 12358212]
17. US Preventive Services Task Force. Screening for depression in adults: US preventive services task force recommendation statement. Ann Intern Med 2009 Dec 1;151(11):784-792. [doi: 10.7326/0003-4819-151-11-200912010-00006] [Medline: 19949144]
18. Corrigan P. How stigma interferes with mental health care. Am Psychol 2004 Oct;59(7):614-625. [doi: 10.1037/0003-066X.59.7.614] [Medline: 15491256]
19. Beck AT, Steer RA, Brown GK. BDI-II, Beck Depression Inventory: Manual. Second Edition. San Antonio, Texas: Psychological Corporation; 1996.
20. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the patient health questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007 Nov;22(11):1596-1602 [FREE Full text] [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]
21. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
22. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell scale for depression in dementia. Biol Psychiatry 1988 Feb 1;23(3):271-284. [doi: 10.1016/0006-3223(88)90038-8] [Medline: 3337862]
23. Williams JB. Standardizing the Hamilton depression rating scale: past, present, and future. Eur Arch Psychiatry Clin Neurosci 2001;251(Suppl 2):II6-I12. [doi: 10.1007/BF03035120] [Medline: 11824839]
24. Valenstein M, Vijan S, Zeber JE, Boehm K, Buttar A. The cost-utility of screening for depression in primary care. Ann Intern Med 2001 Mar 6;134(5):345-360. [doi: 10.7326/0003-4819-134-5-200103060-00007] [Medline: 11242495]
25. Kerr LK, Kerr Jr LD. Screening tools for depression in primary care: the effects of culture, gender, and somatic symptoms on the detection of depression. West J Med 2001 Nov;175(5):349-352 [FREE Full text] [doi: 10.1136/ewjm.175.5.349] [Medline: 11694495]
26. Mojtabai R, Olfson M. Proportion of antidepressants prescribed without a psychiatric diagnosis is growing. Health Aff (Millwood) 2011 Aug;30(8):1434-1442. [doi: 10.1377/hlthaff.2010.1024] [Medline: 21821561]
27. Mojtabai R. Clinician-identified depression in community settings: concordance with structured-interview diagnoses. Psychother Psychosom 2013;82(3):161-169. [doi: 10.1159/000345968] [Medline: 23548817]
28. Dowrick C, Frances A. Medicalising unhappiness: new classification of depression risks more patients being put on drug treatment from which they will not benefit. Br Med J 2013 Dec 9;347:f7140. [doi: 10.1136/bmj.f7140] [Medline: 24322400]
29. Thombs BD, Ziegelstein RC, Roseman M, Kloda LA, Ioannidis JP. There are no randomized controlled trials that support the United States Preventive Services Task Force Guideline on screening for depression in primary care: a systematic review. BMC Med 2014 Jan 28;12:13 [FREE Full text] [doi: 10.1186/1741-7015-12-13] [Medline: 24472580]
30. Thombs BD, Coyne JC, Cuijpers P, de Jonge P, Gilbody S, Ioannidis JP, et al. Rethinking recommendations for screening for depression in primary care. CMAJ 2012 Mar 6;184(4):413-418 [FREE Full text] [doi: 10.1503/cmaj.111035] [Medline: 21930744]
31. Gilbody S, Sheldon T, Wessely S. Should we screen for depression? Br Med J 2006 Apr 29;332(7548):1027-1030 [FREE Full text] [doi: 10.1136/bmj.332.7548.1027] [Medline: 16644833]
32. Charlson ME, Charlson RE, Peterson JC, Marinopoulos SS, Briggs WM, Hollenberg JP. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. J Clin Epidemiol 2008 Dec;61(12):1234-1240. [doi: 10.1016/j.jclinepi.2008.01.006] [Medline: 18619805]

33. DeJean D, Giacomini M, Vanstone M, Brundisini F. Patient experiences of depression and anxiety with chronic disease: a systematic review and qualitative meta-synthesis. Ont Health Technol Assess Ser 2013;13(16):1-33 [FREE Full text] [Medline: 24228079]
34. Kasthurirathne SN. IUPUI Scholar Works Repository. 2018. The Use of Clinical, Behavioral, and Social Determinants of Health to Improve Identification of Patients in Need of Advanced Care for Depression URL: https://scholarworks.iupui.edu/bitstream/handle/1805/17765/Kasthurirathne_iupui_0104D_10328.pdf?sequence=1&isAllowed=y
35. National Library of Medicine - National Institutes of Health. 2016. Unified Medical Language System® (UMLS®): Metathesaurus URL: https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/ [accessed 2018-02-01]
36. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
37. National Library of Medicine - National Institutes of Health. 2015. Fact Sheet: Unified Medical Language System URL: https://www.nlm.nih.gov/pubs/factsheets/umls.html [accessed 2018-01-01]
38. McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, INPC Management Committee. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. Health Aff (Millwood) 2005;24(5):1214-1220. [doi: 10.1377/hlthaff.24.5.1214] [Medline: 16162565]
39. Overhage JM. The Indiana health information exchange. In: Dixon BE, editor. Health Information Exchange: Navigating and Managing a Network of Health Information Systems. Waltham, MA: Academic Press; 2016:267-279.
40. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
41. Kasthurirathne SN, Dixon BE, Gichoya J, Xu H, Xia Y, Mamlin B, et al. Toward better public health reporting using existing off the shelf approaches: a comparison of alternative cancer detection approaches using plaintext medical data and non-dictionary based feature selection. J Biomed Inform 2016 Apr;60:145-152 [FREE Full text] [doi: 10.1016/j.jbi.2016.01.008] [Medline: 26826453]
42. Kasthurirathne SN, Vest JR, Menachemi N, Halverson PK, Grannis SJ. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. J Am Med Inform Assoc 2018 Dec 1;25(1):47-53. [doi: 10.1093/jamia/ocx130] [Medline: 29177457]
43. Qi Y. Random forest for bioinformatics. In: Zhang C, Ma Y, editors. Ensemble Machine Learning: Methods and Applications. New York: Springer; 2012:307-323.
44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
45. Youden WJ. Index for rating diagnostic tests. Cancer 1950 Jan;3(1):32-35. [doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3] [Medline: 15405679]
46. Polani D. Kullback-leibler divergence. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H, editors. Encyclopedia of Systems Biology. New York: Springer; 2013:1087-1088.
47. Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak 2011 Jul 29;11:51 [FREE Full text] [doi: 10.1186/1472-6947-11-51] [Medline: 21801360]
48. Zong W, Huang GB, Chen YJ. Weighted extreme learning machine for imbalance learning. Neurocomputing 2013 Feb;101:229-242. [doi: 10.1016/j.neucom.2012.08.010]
49. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res 2017;18(1):559-563 [FREE Full text]
50. Loeb DF, Binswanger IA, Candrian C, Bayliss EA. Primary care physician insights into a typology of the complex patient in primary care. Ann Fam Med 2015 Sep;13(5):451-455 [FREE Full text] [doi: 10.1370/afm.1840] [Medline: 26371266]
51. Jackson GL, Powers BJ, Chatterjee R, Bettger JP, Kemper AR, Hasselblad V, et al. The patient centered medical home. A systematic review. Ann Intern Med 2013 Dec 5;158(3):169-178. [doi: 10.7326/0003-4819-158-3-201302050-00579] [Medline: 24779044]
52. Kharrazi H, Weiner JP. IT-enabled community health interventions: challenges, opportunities, and future directions. EGEMS (Wash DC) 2014;2(3):1117 [FREE Full text] [doi: 10.13063/2327-9214.1117] [Medline: 25848627]
53. Kharrazi H, Lasser EC, Yasnoff WA, Loonsk J, Advani A, Lehmann HP, et al. A proposed national research and development agenda for population health informatics: summary recommendations from a national expert workshop. J Am Med Inform Assoc 2017 Dec;24(1):2-12 [FREE Full text] [doi: 10.1093/jamia/ocv210] [Medline: 27018264]
54. Cape J, Whittington C, Buszewicz M, Wallace P, Underwood L. Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. BMC Med 2010 Jun 25;8:38 [FREE Full text] [doi: 10.1186/1741-7015-8-38] [Medline: 20579335]
55. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. Br Med J 2015;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26759375]
56. Katon WJ. Epidemiology and treatment of depression in patients with chronic medical illness. Dialogues Clin Neurosci 2011;13(1):7-23 [FREE Full text] [Medline: 21485743]

57. World Health Organization. 2014. Social Determinants of Mental Health URL: https://www.who.int/mental_health/publications/gulbenkian_paper_social_determinants_of_mental_health/en/
58. Fryers T, Melzer D, Jenkins R, Brugha T. The distribution of the common mental disorders: social inequalities in Europe. Clin Pract Epidemiol Ment Health 2005 Sep 5;1:14 [FREE Full text] [doi: 10.1186/1745-0179-1-14] [Medline: 16143042]
59. Kharrazi H, Chi W, Chang HY, Richards TM, Gallagher JM, Knudson SM, et al. Comparing population-based risk-stratification model performance using demographic, diagnosis and medication data extracted from outpatient electronic health records versus administrative claims. Med Care 2017 Dec;55(8):789-796. [doi: 10.1097/MLR.0000000000000754] [Medline: 28598890]
60. Kharrazi H, Anzaldi LJ, Hernandez L, Davison A, Boyd CM, Leff B, et al. The value of unstructured electronic health record data in geriatric syndrome case identification. J Am Geriatr Soc 2018 Aug;66(8):1499-1507. [doi: 10.1111/jgs.15411] [Medline: 29972595]
61. Haykin S. Neural Networks: A Comprehensive Foundation. New York: Pearson; 2004.
62. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013 Aug;35(8):1798-1828. [doi: 10.1109/TPAMI.2013.50] [Medline: 23787338]

## Abbreviations

**AUC:** area under the curve
**ICD:** International Classification of Diseases
**INPC:** Indiana Network for Patient Care
**PHQ:** Patient Health Questionnaire
**UMLS:** Unified Medical Language System

*Edited by G Eysenbach; submitted 25.02.19; peer-reviewed by A Flynn, H Kharrazi; comments to author 11.04.19; revised version received 01.06.19; accepted 12.06.19; published 22.07.19.*

*Please cite as:*
*Kasthurirathne SN, Biondich PG, Grannis SJ, Purkayastha S, Vest JR, Jones JF*
*Identification of Patients in Need of Advanced Care for Depression Using Data Extracted From a Statewide Health Information Exchange: A Machine Learning Approach*
*J Med Internet Res 2019;21(7):e13809*
*URL: http://www.jmir.org/2019/7/e13809/*
*doi:10.2196/13809*
*PMID:31333196*

Original Paper

# Identifying Key Target Audiences for Public Health Campaigns: Leveraging Machine Learning in the Case of Hookah Tobacco Smoking

Kar-Hai Chu[1*], PhD; Jason Colditz[1*], MEd; Momin Malik[2*], MS; Tabitha Yates[1*]; Brian Primack[1*], MD, PhD

[1]School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[2]Carnegie Mellon University, Pittsburgh, PA, United States

[*]all authors contributed equally

**Corresponding Author:**
Kar-Hai Chu, PhD
School of Medicine
University of Pittsburgh
230 McKee Place
Suite 600
Pittsburgh, PA, 15213
United States
Phone: 1 4126922578
Email: chuk@pitt.edu

## Abstract

**Background:** Hookah tobacco smoking (HTS) is a particularly important issue for public health professionals to address owing to its prevalence and deleterious health effects. Social media sites can be a valuable tool for public health officials to conduct informational health campaigns. Current social media platforms provide researchers with opportunities to better identify and target specific audiences and even individuals. However, we are not aware of systematic research attempting to identify audiences with mixed or ambivalent views toward HTS.

**Objective:** The objective of this study was to (1) confirm previous research showing positively skewed HTS sentiment on Twitter using a larger dataset by leveraging machine learning techniques and (2) systematically identify individuals who exhibit mixed opinions about HTS via the Twitter platform and therefore represent key audiences for intervention.

**Methods:** We prospectively collected tweets related to HTS from January to June 2016. We double-coded sentiment for a subset of approximately 5000 randomly sampled tweets for sentiment toward HTS and used these data to train a machine learning classifier to assess the remaining approximately 556,000 HTS-related Twitter posts. Natural language processing software was used to extract linguistic *features* (ie, language-based covariates). The data were processed by machine learning tools and algorithms using R. Finally, we used the results to identify individuals who, because they had consistently posted both positive and negative content, might be ambivalent toward HTS and represent an ideal audience for intervention.

**Results:** There were 561,960 HTS-related tweets: 373,911 were classified as positive and 183,139 were classified as negative. A set of 12,861 users met a priori criteria indicating that they posted both positive and negative tweets about HTS.

**Conclusions:** Sentiment analysis can allow researchers to identify audience segments on social media that demonstrate ambiguity toward key public health issues, such as HTS, and therefore represent ideal populations for intervention. Using large social media datasets can help public health officials to preemptively identify specific audience segments that would be most receptive to targeted campaigns.

***(J Med Internet Res 2019;21(7):e12443)*** doi:10.2196/12443

**KEYWORDS**

smoking water pipes; waterpipe tobacco; tobacco; smoking; social media; public health; infodemiology; infoveillance; machine learning

## Introduction

Hookah tobacco smoking (HTS)—also called waterpipe, shisha, or narghile—has increased substantially in popularity [1]. HTS is common among college students, with ever-use rates ranging from 30% to 46% [2,3]. It has been associated with multiple health conditions including cancer, cardiovascular disease, decreased pulmonary function, and nicotine dependence [4-6]. Owing to its high prevalence and deleterious health effects, HTS is a particularly important issue for public health professionals to address.

Social media sites can be a valuable tool for public health officials to conduct informational health campaigns. This process can be informed by machine learning approaches that are able to conduct topic classification [7] or sentiment analysis [8] in very large datasets. For example, researchers have also found a large amount of posted content on Twitter that describes HTS as positive [9] and normalizes the activity [10]. This is concerning because social media exposure to tobacco products is known to influence attitudes and future smoking behavior [11,12]. In response, public health departments have leveraged social media sites to conduct informational campaigns around HTS, including the Center for Disease Control and Prevention's Tips from Former Smokers [13,14] and *ShishAware* [15].

Although programs such as these tend to use broad approaches, it may be more advantageous to tailor HTS-related educational messages to targeted groups or individuals [16]. Current social media platforms provide researchers with opportunities to better identify and target specific audiences and even individuals. However, we are not aware of systematic research attempting to identify audiences with mixed or ambivalent views toward HTS.

Therefore, this study is designed to accomplish 2 aims: (1) confirm previous research in HTS sentiment on Twitter [9,10] using a larger dataset by leveraging machine learning techniques and (2) systematically identify individuals who exhibit mixed opinions toward HTS via the Twitter platform. The latter procedure can provide actionable data for public health officials in developing educational campaigns for wide and efficient dissemination.

## Methods

### Data

Twitter is a microblog platform on which users post *tweets* that are shared either publicly or to their private network of followers. We collected 561,960 HTS-related tweets over 6 months, from January 1 to June 30, 2016. The search terms used were *hookah*, *hookahs*, *hooka*, *shisha*, *sheesha*, and *narghile*. We do not suggest that these 6 terms represent all possible HTS posted content; however, they follow previous research protocols [17] and allowed for collection of a large dataset of HTS tweets for our purpose. Data were collected by establishing connections to Twitter's application programming interface, which permits external software to request data. All data collected were publicly available, that is, anyone with an internet connection was able to view the tweet at the time it was collected. Data included full text of the tweet content (maximum of 140 characters) as well as the identifier (ID) of the originating Twitter user to track individual users' tweets over time. Tweet content was obtained in plain text format and reformatted by replacing *emoji* (images and symbols embedded within text) with human-readable counterparts (eg, a heart symbol becomes *[heart emoji]*) [17]. We also replaced specific hyperlinks and usernames with generic placeholders (eg, [URL]; @[USERNAME]). User IDs were then recoded to de-identified numeric IDs (eg, User 1). This study has been reviewed and approved by the Institutional Review Board at the authors' university.

### Procedures

We applied machine learning algorithms to conduct sentiment analysis on HTS content posted to Twitter. Supervised machine learning allows for a relatively small amount of human-coded data to train computerized algorithms that can automatically categorize additional data on a scale that would not be feasible otherwise. We conducted the sentiment analysis in 2 phases. First, a set of approximately 5000 tweets were randomly sampled and categorized as being positive, negative, or both, and if it was commercial by 2 trained coders with experience in categorizing tobacco-related data on Twitter [17]. Inter-rater reliability measures for the human coders were all substantial or better [18], with Cohen kappa=.78 for positive/not positive and kappa=.75 for negative/not negative, and kappa=.82 for commercial. The coded tweets were subsequently used as training data for a classifier that automatically coded the sentiment in the remaining approximately 556,000 HTS-related Twitter posts. Natural language processing software RWeka and tm were used to extract linguistic *features* (ie, language-based covariates). The data were processed by machine learning tools in RTextTools.

In the second phase, human coders—informed by the previous classifications—identified Twitter accounts that had posted both positive and negative sentiment tweets about HTS, signaling a potential group of users with mixed or ambivalent views.

### Measures

Each tweet in the training dataset was given 3 categorical codes by 2 coders: (1) positive or not positive, (2) negative or not negative, and (3) commercial or not commercial. Commercial content was defined as anything promoting the sale of a particular hookah product, establishment, or related service (eg, a hookah bar promoting happy hour specials). These tweets were identified based on textual content rather than by the type of Twitter user posting the content (eg, a hookah bar could also post noncommercial content in other contexts; an individual user could promote a hookah bar). This allowed for us to maintain tweet content as the primary unit of sentiment analysis, rather than including user-level metadata (eg, establishment name or profile image) that our text-based machine learning classifier would be unable to judge. Sentiment (ie, positive and negative) was defined as *positive/negative toward HTS* rather than an overall positive or negative expression. This provided an advantage for supervised machine learning, as it allowed the classifier to be content specific, rather than depend on general sentiment terms. As previous research found that HTS-related content skewed positive [9,10], positive tweets were undersampled in the final training data to match the number of coded negative tweets.

### Analysis

To reduce the complexity of machine learning classifications, commercial tweets were included as positive or pro-HTS. After the machine learning completed classification of the approximately 556,000 tweets, 1000 tweets were randomly sampled to calculate 3 performance metrics: (1) precision, calculated as true positives divided by total instances labeled as positive; (2) recall, identical to sensitivity, calculated as true positives divided by total positive instances; and (3) F-score, a weighted average of precision and recall.

Finally, a qualitative analysis of Twitter users who had posted both positive and negative content was conducted. For this content search, tweets were grouped by user; based on an exploratory view by 2 authors of tweets in other topics (ie, beyond hookah-related posts), we decided that those with more than 5 times as many positive as negative posts, and vice versa, would not likely have truly mixed or ambivalent opinions and were removed from consideration. Coders were then tasked to identify potential users who had posted both positive and negative sentiment tweets.

## Results

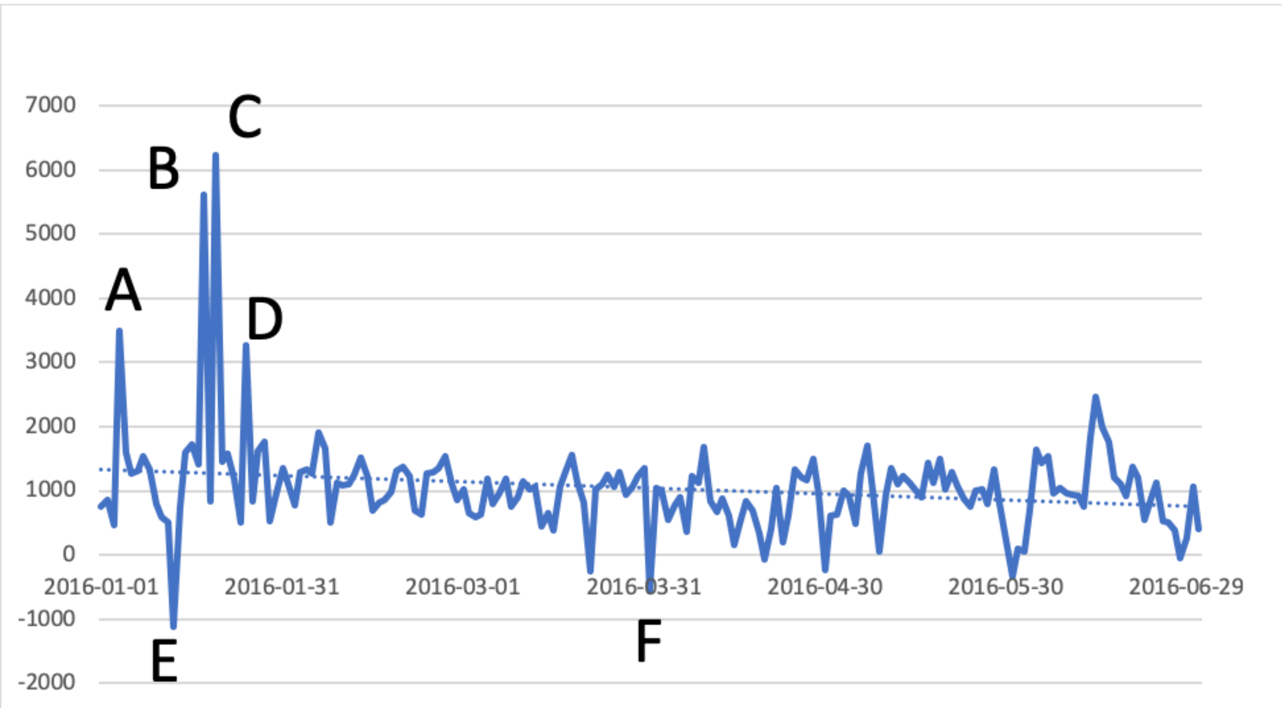
Table 1 provides the results of the full classifications, including summaries of precision and recall. Figure 1 shows the sentiment of HTS tweets over a 6-month period. Daily sentiment scores were calculated by subtracting the number of negative tweets from the number of positive tweets. These ranged from −1116 to 6239, with a mean of 1042 (SD 749).

There were 6 spikes during the 6-month period, defined as days where sentiment increased or decreased by more than 2 standard deviations. These are labeled A-F in Figure 1. The 4 positive spikes were as follows: (1) on January 4, a popular song about HTS (A); (2) on January 18, discussion of an HTS lounge (B); (3) on January 20, description of a concealable HTS device (C); and (4) on January 25, same as point 3 (D). The 2 negative instances were as follows: (1) on January 13, a report on the high levels of tar when using HTS (E) and (2) on April 1, an athlete was *caught* smoking from a hookah (F).

**Table 1.** Results of machine learning classifier with precision/recall metrics (January-June 2016).

| Sentiment | Classified (N=561,960), n (%) | Precision | Recall | F-score |
|---|---|---|---|---|
| Positive | 373,911 (66.53) | 0.92 | 0.81 | 0.86 |
| Negative | 183,139 (32.59) | 0.59 | 0.79 | 0.67 |

**Figure 1.** Sentiment of hookah tobacco smoking tweets over a 6-month period.



A total of 291,602 users posted HTS content over the period of the study, ranging from 1 to 6501 tweets each, with a median of 1. To identify users considered having ambivalent or mixed opinions about HTS, the criterion was defined as someone who had posted at least one positive and one negative tweet over the observed period. When we removed users that posted content with either positive or negative sentiment at a ratio greater than 5:1, 4.41% of users (12,861/291,602) remained. Of these users, we randomly sampled 1.00% of users (129/12,861) and selected all of their tweets to be qualitatively examined by 2 coders. There were 37 (29%) users classified as having clear ambivalent or mixed opinions about HTS. Examples of these users and their tweets are displayed in Table 2.

**Table 2.** A sample of 10 users (out of 37) who posted both positive and negative tweets about hookah tobacco smoking on Twitter between January and June 2016.

| User | Tweet |
|---|---|
| 1 | • Wednesday about to be lit lmao I need a hookah man[a]<br>• I don't want hookah no more dawg lmao |
| 2 | • Life feels so good when you are smoking hookah.. [blushing emoji][a]<br>• SO I TRIED VAPING TODAY, ON 108Hz.. HAHAHA fucking hard but such thick clouds, vaping is the best! gotta quit shisha and start vaping now! |
| 3 | • @[USERNAME] stop smokin hookah then<br>• The hookah spot was rockin wit bitches feenin for cancer smh[a] |
| 4 | • I'm smoking hookah in front of my building right now [URL][a]<br>• My goal is to not DJ any spots with Hookah this summer |
| 5 | • Almost all my male friends love hookah smh[a]<br>• Trying to put plans together for Chandra's birthday and I have to make sure hookah is involved [weary emoji] |
| 6 | • Man y'all be paying 20 dollars at hookah spots to stare at each other [sobbing emoji][a]<br>• I only smoke shisha once in a while tbh lmao and wth we got jobs and school [URL] |
| 7 | • She was sent from the heavens... She don't smoke hookah or know about lemonade. #Skinny<br>• I gotta find a way to make crab flavored hookah tobacco. #Skinny[a] |
| 8 | • FAM be proud of me I havent smoked hookah ALL year -@[USERNAME]<br>• My ramadan nights bouta consist of me sitting on the porch till 5am skyping and smoking hookah.[a] |
| 9 | • I wish hookah never existed [URL]<br>• There's no hookah so why go [URL][a] |
| 10 | • I've done hookah less than 5 times[a]<br>• whenever I smoke hookah I wanna throw up |

[a]Positive tweets.

## Discussion

### Principal Findings

This study combined several lines of research to integrate machine learning and online social media to inform public health research. We completed a 6-month sentiment analysis of HTS posts on Twitter and found that a majority (67%) of content posted about HTS were positive. This confirmed previous studies of HTS sentiment on Twitter with smaller datasets [9,10]. The 6 spikes in Figure 1 offered insight into how surveillance of HTS sentiment over time might be informative to public health departments. First, the identification of a positive spike following a newly released song referencing hookahs and the negative spike after a report of an athlete using a hookah demonstrated a capable process of having instantaneous access to public opinions of real-world events related to HTS. Immediate responses could be developed to address or leverage these situations, depending on the sentiment of the message. Second, reports of new HTS devices or technologies can also be quickly identified and allow for fast countermeasures to be taken as needed. Third, although the discussion of an HTS lounge also caused a positive spike, it was most likely to spread through use of automated bots or nonhuman accounts. Previous literature has found a significant number of Twitter accounts that discuss electronic cigarettes (e-cigarettes)—especially for advertisements—stemming from automated accounts [19]; it is possible a similarly high percentage are used for HTS as well. Additional research is needed to determine the prevalence of automated accounts that discuss HTS, as well as what effect this might have on the messages that are being spread.

Extending beyond a descriptive analysis of Twitter sentiment for surveillance, the second phase successfully identified users posting mixed or ambivalent sentiment tweets about HTS. The realization of our strategy was substantiated by 2 coders' ability to detect several clear examples of posted tweets about HTS that differed in sentiment. We limited our discussion to 10 users for the sake of brevity, although more were discovered. These examples (Table 2) included people who had only tried HTS a limited number of times (users 6 and 10), people against HTS but who go to HTS lounges (user 3), people with lots of friends who use hookah (user 5), people trying to avoid or quit HTS (users 1, 2, 4, and 8), and people who recognize the potential negative side effects of HTS but still support or use it (users 7 and 9). These Twitter users could be an ideal audience for any public health campaign that is focused on HTS prevention or cessation. As the examples demonstrated, there was no clear pattern of vocabulary, topic, or other semantic features that were obvious in the data; instead, it was only by applying our method of aggregating tweets by users and then searching for mixed-sentiment content—from a dataset of approximately 561,000 posts—that we were able to identify these Twitter users. The ability to capture all of these users established the strength of the technique, as finding these users could not be easily accomplished manually.

Concerns exist when using machine learning to analyze topics with skewed data distributions; imbalanced data can reduce a machine learning algorithm's capacity to properly classify data that are disproportionately small, also called a minority class. This is due to conventional algorithms being biased toward the majority class in an effort to optimize error rates [20]. For example, if a dataset of 100 tweets contained 90 positives and 10 negatives, a classifier that correctly labeled 83 positive tweets and 6 negative tweets would have an 89% accuracy; however, another classifier that labeled every tweet as positive would seem to score a *better* 90%, even though it was unable to detect any negative content. As previous research has shown that HTS-related content on social media tended to strongly be skewed positive, we chose a strategy to undersample positive tweets [21], trading decreased sensitivity for increased specificity with regard to the minority class. Similar results to previous studies combined with a reasonably high negative recall (Table 1) suggest that our approach was successful.

### Limitations

Limitations of our study include only using publicly available data from Twitter; inclusion of private Twitter content might lead to different results. Twitter user demographics can also limit the generalizability of these data. There is a possibility that a small percentage of HTS-related tweets are actually discussing using a hookah to smoke marijuana, although none were found in the sample that was human-coded. As we focused on linguistic features of the text, other media sources such as images or videos were not analyzed; expanding beyond linguistic features might also help improve the lower negative post precision. In addition, our strategy in choosing supervised machine learning restricts the results to HTS; a classifier would need to be retrained with content-specific data for other uses.

### Comparison With Previous Work

In recent years, public health officials have developed Twitter campaigns addressing tobacco products such as e-cigarettes. However, these campaigns can be hijacked by opposing organizations and result in countercampaigns. For example, the Chicago Department of Public Health released a series of messages about e-cigarettes a week before a scheduled vote on local regulation by the city council. Unfortunately, hundreds of tweets responded with opposing claims such as the health department was lying or disseminating propaganda [22]. In a similar fashion, the California Department of Public Health launched an anti–e-cigarette media campaign on Twitter called *Still Blowing Smoke*. As with the case in Chicago, a countercampaign quickly launched, called *Not Blowing Smoke*, that gained more attention [23]. In both cases, the original topics were meant to be general and far reaching; regrettably, those properties allowed for opposing organizations to have a single challenging message that could be disseminated quickly to the same population. Our study provides a unique method that generates data to identify precise subpopulations and specific topics that organically emerged from natural discourse. Having identified users with mixed opinions about HTS can be used as actionable data for a public health campaign. Rather than depend solely on expert knowledge to develop a single campaign that runs the risk of being misused by an opposing organization, this method provides the empirical evidence to build informational campaigns grounded on information that is actually needed by HTS users or groups.

### Conclusions

Tobacco control researchers focused on HTS should endeavor to develop campaigns that target this audience segment. Twitter has been proposed as a monitoring setting for public health [24,25], but the mixed success of other efforts to use mass user-generated data for large-scale public health detection [26] reveals the nonrepresentative nature of online and social media data in demographics and user patterns [27]. However, using social media as a setting to identify and reach potentially receptive audiences helps to avoid these threats to external validity. This mixed-sentiment strategy is not tobacco-specific and could be implemented in any public health setting. The approach could significantly improve the efficiency of any health campaign, allowing public health departments to be mindful with available resources and be confident of higher success rates.

Public health campaigns have frequently used mass media to disseminate educational or informational messages. State-level public health departments have also utilized Twitter to conduct informational tobacco campaigns with mixed results [22,23]. Although steps have been taken to leverage social media for these endeavors, officials have used strategies that do not always incorporate new types of information that technology can provide. We have demonstrated that user sentiment around HTS can and does change over time. Thus, it may be worthwhile to target public health interventions to individuals expressing positive or neutral sentiment toward HTS. Applying techniques in machine learning on large social media datasets can help public health officials to preemptively identify specific audience segments that would be most receptive to targeted campaigns. This allows a more purposeful and efficient method of producing changes in opinions, beliefs, or behaviors.

## Acknowledgments

This study is supported through grants from the Agency for Healthcare Research and Quality (K12HS022989) and National Cancer Institute (R01CA225773; K07CA222338). The technical infrastructure for this study was supported by National Science Foundation award number ACI-1445606 at the Pittsburgh Supercomputing Center. The funders had no involvement in review, approval, or decision to publish this manuscript.

## Conflicts of Interest

None declared.

## References

1. Centers for Disease Control and Prevention. 2016. Hookahs URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/tobacco_industry/hookahs/index.htm
2. Barnett TE, Smith T, He Y, Soule EK, Curbow BA, Tomar SL, et al. Evidence of emerging hookah use among university students: a cross-sectional comparison between hookah and cigarette use. BMC Public Health 2013 Apr 5;13:302 [FREE Full text] [doi: 10.1186/1471-2458-13-302] [Medline: 23560649]
3. Primack BA, Shensa A, Kim KH, Carroll MV, Hoban MT, Leino EV, et al. Waterpipe smoking among US university students. Nicotine Tob Res 2013;15(1):29-35 [FREE Full text] [doi: 10.1093/ntr/nts076] [Medline: 22641433]
4. Haddad L, Kelly DL, Weglicki LS, Barnett TE, Ferrell AV, Ghadban R. A systematic review of effects of waterpipe smoking on cardiovascular and respiratory health outcomes. Tob Use Insights 2016;9:13-28 [FREE Full text] [doi: 10.4137/TUI.S39873] [Medline: 27398028]
5. Kadhum M, Sweidan A, Jaffery AE, Al-Saadi A, Madden B. A review of the health effects of smoking shisha. Clin Med (Lond) 2015 Jun;15(3):263-266 [FREE Full text] [doi: 10.7861/clinmedicine.15-3-263] [Medline: 26031977]
6. Sidani JE, Shensa A, Shiffman S, Switzer GE, Primack BA. Behavioral associations with waterpipe tobacco smoking dependence among US young adults. Addiction 2016 Feb;111(2):351-359 [FREE Full text] [doi: 10.1111/add.13163] [Medline: 26417942]
7. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. J Med Internet Res 2015 Aug 25;17(8):e208 [FREE Full text] [doi: 10.2196/jmir.4392] [Medline: 26307512]
8. Myslín M, Zhu S, Chapman W, Conway M. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174. [doi: 10.2196/jmir.2534] [Medline: 23989137]
9. Grant A, O'Mahoney H. Portrayal of waterpipe (shisha, hookah, nargile) smoking on Twitter: a qualitative exploration. Public Health 2016 Nov;140:128-135. [doi: 10.1016/j.puhe.2016.07.007] [Medline: 27520707]
10. Krauss MJ, Sowles SJ, Moreno MA, Zewdie K, Grucza RA, Bierut LJ, et al. Hookah-related Twitter chatter: a content analysis. Prev Chronic Dis 2015 Jul 30;12:E121 [FREE Full text] [doi: 10.5888/pcd12.150140] [Medline: 26226068]
11. Depue JB, Southwell BG, Betzner AE, Walsh BM. Encoded exposure to tobacco use in social media predicts subsequent smoking behavior. Am J Health Promot 2015;29(4):259-261. [doi: 10.4278/ajhp.130214-ARB-69] [Medline: 24670071]
12. Yoo W, Yang J, Cho E. How social media influence college students' smoking attitudes and intentions. Comput Human Behav 2016 Nov;64:173-182 [FREE Full text] [doi: 10.1016/j.chb.2016.06.061] [Medline: 27956757]
13. Centers for Disease Control and Prevention. 2018. Tips From Former Smokers URL: https://www.cdc.gov/tobacco/campaign/tips/index.html
14. Chung JE. A smoking cessation campaign on Twitter: understanding the use of Twitter and identifying major players in a health campaign. J Health Commun 2016;21(5):517-526. [doi: 10.1080/10810730.2015.1103332] [Medline: 27115179]
15. Jawad M, Abass J, Hariri A, Akl EA. Social media use for public health campaigning in a low resource setting: the case of waterpipe tobacco smoking. Biomed Res Int 2015;2015:562586 [FREE Full text] [doi: 10.1155/2015/562586] [Medline: 26273631]
16. Thackeray R, Neiger BL, Keller H. Integrating social media and social marketing: a four-step process. Health Promot Pract 2012 Mar;13(2):165-168. [doi: 10.1177/1524839911432009] [Medline: 22382492]
17. Colditz JB, Chu K, Emery SL, Larkin CR, James AE, Welling J, et al. Toward real-time infoveillance of Twitter health messages. Am J Public Health 2018 Aug;108(8):1009-1014. [doi: 10.2105/AJPH.2018.304497] [Medline: 29927648]
18. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37-46. [doi: 10.1177/001316446002000104]
19. Allem JP, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. J Med Internet Res 2016 Dec 9;18(8):e219 [FREE Full text] [doi: 10.2196/jmir.6185] [Medline: 27507563]

20. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng 2009;21(9):1263-1284. [doi: 10.1109/TKDE.2008.239]
21. Wallace BC, Small K, Brodley CE, Trikalinos TA. Class Imbalance, Redux. In: Proceedings of the 2011 IEEE 11th International Conference on Data Mining. 2011 Presented at: ICDM'11; December 11-14, 2011; Vancouver, Canada p. 754-763. [doi: 10.1109/ICDM.2011.33]
22. 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 Oct 16;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
23. Allem JP, Escobedo P, Chu KH, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. Tob Control 2017;26(2):226-229 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052757] [Medline: 26956467]
24. Lampos V, de Bie T, Cristianini N. Flu detector-tracking epidemics on Twitter. In: Balcázar JL, Bonchi F, Gionis A, Sebag M, editors. Machine Learning and Knowledge Discovery in Databases. Berlin: Springer; 2010:599-602.
25. Sadilek A, Kautz H, Silenzio V. Predicting Disease Transmission From Geo-Tagged Micro-Blog Data. In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. 2012 Presented at: AAAI'12; July 22-26, 2012; Toronto, Ontario p. 136-142.
26. 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]
27. Blank G. The digital divide among Twitter users and its implications for social research. Soc Sci Comput Rev 2016 Oct 12;35(6):679-697. [doi: 10.1177/0894439316671698]


## Abbreviations

**e-cigarette:** electronic cigarette
**HTS:** hookah tobacco smoking
**ID:** identifier


*Edited by G Eysenbach; submitted 07.10.18; peer-reviewed by JE Chung, Y Guan; comments to author 03.02.19; revised version received 31.03.19; accepted 20.05.19; published 08.07.19.*

*Please cite as:*
*Chu KH, Colditz J, Malik M, Yates T, Primack B*
*Identifying Key Target Audiences for Public Health Campaigns: Leveraging Machine Learning in the Case of Hookah Tobacco Smoking*
*J Med Internet Res 2019;21(7):e12443*
*URL: http://www.jmir.org/2019/7/e12443/*
*doi:10.2196/12443*
*PMID:31287063*

Original Paper

# Identification and Quantification of Gaps in Access to Autism Resources in the United States: An Infodemiological Study

Michael Ning[1,2], BS; Jena Daniels[1,2], BS; Jessey Schwartz[1,2], BA; Kaitlyn Dunlap[1,2], BS, MRES; Peter Washington[1,2], BA, MS; Haik Kalantarian[1,2], BS, MS, PhD; Michael Du[1,2], BA; Dennis P Wall[1,2], BS, MSCE, PhD

[1]Division of Systems Medicine, Department of Pediatrics, Stanford University, Stanford, CA, United States

[2]Department of Biomedical Data Science, Stanford University, Stanford, CA, United States

**Corresponding Author:**
Dennis P Wall, BS, MSCE, PhD
Department of Biomedical Data Science
Stanford University
MSOB Building, X143
1265 Welch Rd
Stanford, CA, 94305
United States
Phone: 1 650 497 9214
Email: dpwall@stanford.edu

## Abstract

**Background:** Autism affects 1 in every 59 children in the United States, according to estimates from the Centers for Disease Control and Prevention's Autism and Developmental Disabilities Monitoring Network in 2018. Although similar rates of autism are reported in rural and urban areas, rural families report greater difficulty in accessing resources. An overwhelming number of families experience long waitlists for diagnostic and therapeutic services.

**Objective:** The objective of this study was to accurately identify gaps in access to autism care using GapMap, a mobile platform that connects families with local resources while continuously collecting up-to-date autism resource epidemiological information.

**Methods:** After being extracted from various databases, resources were deduplicated, validated, and allocated into 7 categories based on the keywords identified on the resource website. The average distance between the individuals from a simulated autism population and the nearest autism resource in our database was calculated for each US county. Resource load, an approximation of demand over supply for diagnostic resources, was calculated for each US county.

**Results:** There are approximately 28,000 US resources validated on the GapMap database, each allocated into 1 or more of the 7 categories. States with the greatest distances to autism resources included Alaska, Nevada, Wyoming, Montana, and Arizona. Of the 7 resource categories, diagnostic resources were the most underrepresented, comprising only 8.83% (2472/28,003) of all resources. Alarmingly, 83.86% (2635/3142) of all US counties lacked any diagnostic resources. States with the highest diagnostic resource load included West Virginia, Kentucky, Maine, Mississippi, and New Mexico.

**Conclusions:** Results from this study demonstrate the sparsity and uneven distribution of diagnostic resources in the United States, which may contribute to the lengthy waitlists and travel distances—barriers to be overcome to be able to receive diagnosis in specific regions. More data are needed on autism diagnosis demand to better quantify resource needs across the United States.

***(J Med Internet Res 2019;21(7):e13094)*** doi:10.2196/13094

**KEYWORDS**

autism; autism spectrum disorder; crowdsourcing; prevalence; resources; infodemiology; epidemiology

## Introduction

### Background

Autism spectrum disorder (ASD), a heterogeneous neurodevelopmental disorder characterized by repetitive and restricted behaviors and interests and social communication impairments, is the fastest growing developmental disorder in the United States [1]. The prevalence of ASD has increased by approximately 700% since 1990 and now affects 1 in 59 children [1-3].

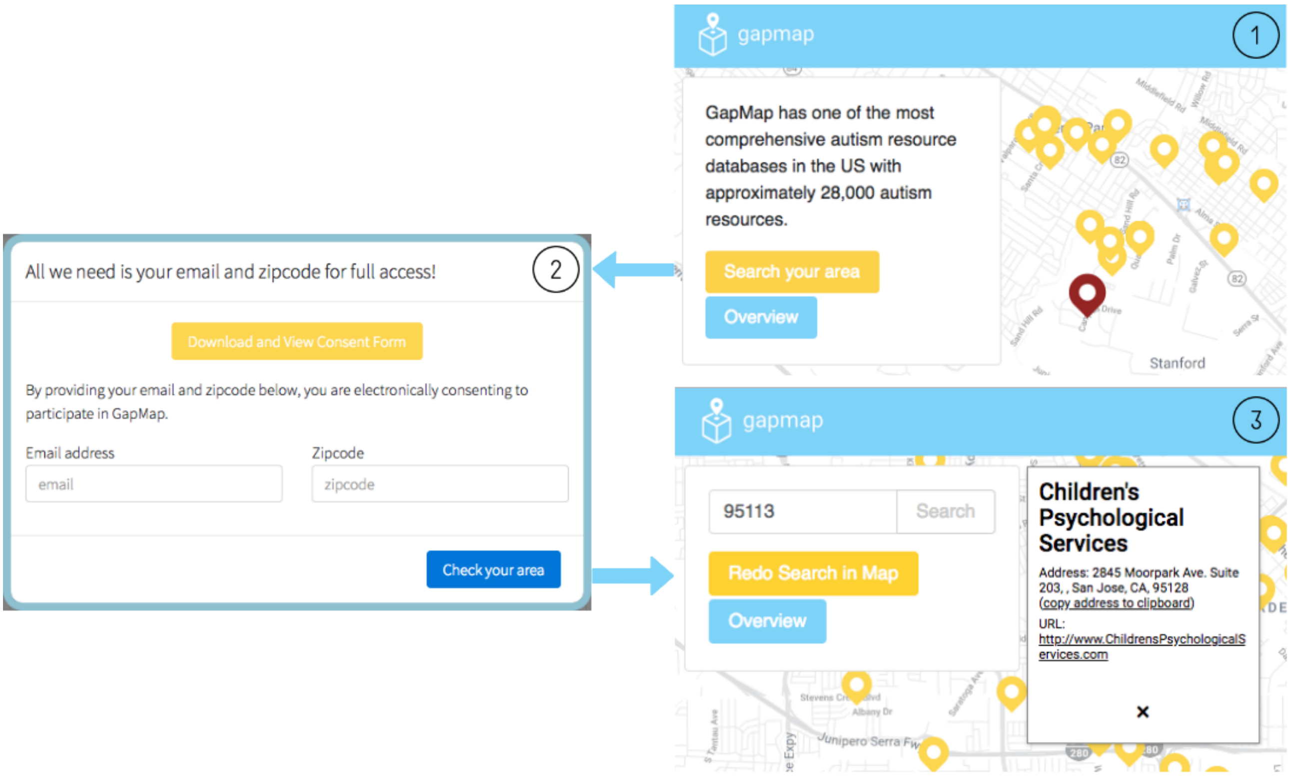
Early access to diagnostic and therapeutic resources is essential for improving outcomes among children with ASD; however, families of individuals with autism often struggle to secure a diagnosis and find adequate therapeutic services [4-6]. The average age of diagnosis in the United States is estimated to be over 4 years [7], and about 27.1% of children remain undiagnosed at 8 years [8], with these figures being even higher among ethnic minorities [9]. Initial concerns are often already noted by families before children reach 2 years, making reliable diagnosis possible given adequate access to diagnostic resources [1,6,10-12]. Unfortunately, clinicians are often overburdened by the increasing population in need of care, creating waitlists exceeding 12 months for diagnosis. Waitlists are even longer for families in rural and underserved areas with a lower socioeconomic status, and an estimated 42% report having to travel to another city for a diagnosis [4,6,13,14]. A staggering 63% of families obtained a diagnosis for their child only after their third visit with a professional [6]. These delays in diagnosis and therapy leave many children untreated until after sensitive periods of development have passed [15-17]. Further complicating matters, treatment for ASD can cost a family up to US $100,000 annually, adding up to US $3.2 million in medical expenses over a lifetime [18,19]. More than half of all families report a lack of information and advice and little opportunity for parents to get involved in developing and reviewing treatment plans for their children [6,20].

## Objective

Although there are indicators of significant resource shortages for families affected by autism in the United States [4-6], few studies have specifically explored autism resource epidemiology in the United States. Quantifying the expanding imbalance between clinicians and families in need of care and identifying the extent of gaps in access to autism resources are both essential steps toward providing support to individuals with autism. To address this need, we developed GapMap [21-23], a centralized crowd-powered Web platform that provides individuals affected by autism with colocated autism resources in their community (Figure 1).

Our preliminary analysis of data derived from GapMap has shown that in the United States, children who live closer to diagnostic centers are more likely to be diagnosed, highlighting a gap in access to care for families in rural communities [21]. This study expands upon these findings with additional data and analyses. In this study, we generated a comprehensive representation of autism resources in the United States, categorized each listed resource by service type, and examined more granular gaps in access in terms of distance and availability. For each of the 3142 counties in the United States, we estimated the average distance between individuals with autism and the nearest autism resource and approximated an annual demand over supply ratio of diagnostic resources.

**Figure 1.** An example of the preliminary mapping interface for GapMap. (1) The landing page shows resources near Stanford University; (2) participants can electronically consent and participate from any desktop or mobile device; (3) after email and zip code submission, resources are shown near the zip code area.



## Methods

### GapMap

GapMap is connected to a MySQL database that stores autism resource metadata including, but not limited to, program names, addresses, service types, and geo-coordinates. GapMap's front-end is written in React.js, and the back-end server runs on Amazon Web Services (AWS) application programming interface (API) Gateway, which connects to AWS Lambda and executes JavaScript packages to communicate with our MySQL database hosted on Amazon Relational Database Service. GapMap allows users to search for resources near a given area and populates a map with markers representing each resource that expands into an information box containing its respective program name, address, and website. Due to its large set of resource geo-coordinates, the GapMap database provides the data needed to perform numerical analyses to further our understanding and characterization of resource epidemiology.

### Resource Database Validation and Categorization

The initial resource database for GapMap [21] was created in 2015 by mining 3 extensive and publicly available autism resource databases (eg, Autism Speaks [24], Autism Source [25], and Parents Helping Parents [26]). Each resource contained the following attributes: *Program Name*, *Program Description*, *Full Address*, *Phone*, *Email*, *URL*, *Latitude*, *Longitude*, and *Category*. Not all resources in GapMap's original database were mappable, relevant, or up to date. To address this issue, a resource was discarded if it contained a post office box or an invalid physical address, was associated with a website that excluded keywords related to autism or was associated with a broken or missing website and lacked additional contact information. A resource with a broken or missing URL was unable to be validated; however, if an email address was provided, the resource was kept in the database for future validation and confirmation.

In an effort to increase comprehensiveness of the GapMap resource and to ensure it remains up to date, we used ParseHub [27], a third-party Web scraper, to extract an updated list of all resources listed on Autism Speaks. In addition, we extracted resources from the Google Places database using Google Places API [28] under the following search parameters: the keyword *autism*, city, state, and search radius. To ensure a comprehensive resource extraction, resources were searched within a 13-km radius of each city center. This radius includes the largest US city (Sitka, Alaska, which has a total area of 12,461 km$^2$). The extracted resources from both databases were merged with the GapMap database, deduplicated based on geo-coordinates and program names, and processed through our validation pipeline.

We assigned resources in the GapMap database to 7 primary resource categories: *Diagnosis*, *Therapy*, *Health*, *Education*, *Recreation*, *Support*, and *Other*. These categories were determined by our clinical coordinators through comparisons between resource categories found on Autism Speaks, Autism Source, and Parents Helping Parents. Resources may include multiple resource categories under the *category* attribute. Table 1 maps the service type attributes already associated with a resource from its native database to 1 of our 7 primary resource categories. Websites associated with resources without a service type were scraped for keywords listed in Table 1 and allocated accordingly. Resources that were not captured by any of the keywords listed in Table 1, for entities such as "...Special Education Attorneys," were allocated to the category *Other*.

**Table 1.** Keywords used to allocate resources to the Diagnosis, Therapy, Health, Education, Recreation, and Support categories on GapMap.

| Keywords | Allocated category |
| --- | --- |
| Diagnosis, Where to get an Autism Diagnosis, Diagnostic, Assessments and Diagnosis | Diagnosis |
| Therapy; Early Intervention: Ages Birth-3; Interventions; Related Services; Early Intervention Services; Augmentative and Alternative Communication; Equine Therapy; Music Therapy; Occupational Therapy; Physical Therapy; Sensory Integration; Social Skills; Speech and Language Therapy; Applied Behavior Analysis (ABA); Floortime or DIR; Other Interventions; Picture Exchange Communication (PEC); Relationship Development Intervention (RDI); SCERTS Model; TEACCH; Verbal Behavior; Early Intervention; Related Services (Therapists); Speech/Language Therapy; Therapists; Art Therapy; instruction/intervention; Executive Function; ABA (Applied Behavior Analysis); Therapeutic; Speech | Therapy |
| Health, Biomedical Interventions; Health Services; Health and Dental Services; Diet or Nutrition; Other Biomedical Interventions; First Responder Resources; Dentists; Family Practitioners; Gastroenterologists; Inpatient Treatment Care Centers; Neurologists; Other Professionals; Pediatricians: Developmental, Pediatricians: General, Psychiatrists, Psychologists, State Mental Health Centers, Blood Draw or Phlebotomists; Community Mental Health Centers; Crisis Intervention Services; Substance Abuse Treatment; Crisis/Crime Victim Services; Dentist; Mental Health Professional; Other Medical Services; Physician; Doctors | Health |
| Education; Preschool Age: Ages 3-5; School Age: Ages 5-22; Post Secondary Education; Schools: Nonpublic (Private); Schools: Residential; Schools: Preschool; Transition to Adult Services; Academic Supports; Private/Non-Public School; Public School System; Job/Vocational; Transition; Living Skills; Schools; Tutoring; Academy; Preschool; Tutor; School; Learn; Teach; Pre-K | Education |
| Recreation; Recreational and Leisure Activities; Day Programs; Recreation and Community Activities; Camps; After-School Programs; Camps and Recreation; Social Activities; After School Care: Children; Recreation; Play | Recreation |
| Support; Support Groups; Community and Support Network; Advocates; Grandparents; Autism Speaks Communities; Local Autism Events; Local Autism Organizations; Military Family Resources; Online Support Groups; Religious Resources; Support Groups; Autism Society Affiliate (Chapter); Parent Training; Conferences; Other Local Organizations; Faith Community Services; Information and Support; Training; Support/Self Help Groups; Parent Support; Advocate-School District; Advocacy; Volunteer | Support |

### Distance Between Resources and Individuals

As GapMap does not yet collect diagnostic information from users, we represented the population distribution of individuals with autism through a simulation. For each of the 3142 counties in the United States, geo-coordinates were generated in a random distribution inside a county bounding box derived from the 2010 US Census Bureau's set of county Tiger/Line shapefiles [29]. The number of geo-coordinates representing the simulated population density was calculated by dividing the US Census Bureau's 2016 County Population Estimate by 59 to normalize for the most recent autism prevalence rate [30]. This amounted to 5,476,742 coordinate points across the United States, where each of these points represents a simulated individual with autism.

All resources in the GapMap database were tagged with their respective geo-coordinates. The Euclidean distance was calculated between each geo-coordinate representing a simulated individual with autism and the geo-coordinate representing the autism resource from the GapMap database closest to the corresponding individual. These distances were used to calculate the average resource distance for each county. The values were then used to rank counties with gaps between autism resources and individuals with autism in descending order. This analysis was repeated for each of the 7 resource categories.

### Resource Load for Diagnostic Resources

Resource distance is correlated with the number of resources and the proportional distribution of resources in the region, but it is not an effective measure of how well resource demands are being met because of the variations in the number of individuals who can be served between resources. As a result, resource load was calculated. The diagnostic resources in the GapMap database comprises resources that provide medical diagnoses of ASD made by pediatricians, neurologists, and psychiatrists as well as diagnoses that occur through psychological assessments and mental health units. Resource load represents how well diagnostic centers can meet the demands of the individuals in a given county annually. Ideally, this value would be a demand over supply ratio of diagnostic services in a given county annually; however, because of the lack of available data on individuals diagnosed and waitlisted for diagnosis for each state, an approximation of this measure was represented by a resource load formula. In addition, as calculating resource load for each diagnostic center rather than region would lead to overestimated values from the potential recounting of individuals within the proximity of multiple diagnostic resources, the previous resource load formula was adjusted to calculate a comparative value that approximates the demand over supply ratio of autism resources. The new resource load below does not exhibit any meaning as a stand-alone value but can be compared with other state resource load values to estimate how well diagnostic resources can meet individual demand.

The resource load *RL_c*, as computed for diagnostic resources for each of the 3142 US counties, is represented by this formula:

*N_c*, the county population, is derived from the US Census Bureau's 2016 County Population Estimate [30]. *R_c* is the number of diagnostic resources in a given county from the GapMap database. For a more intuitive comparison of resource load values, *f* is the ratio selected to normalize by the lowest state resource load, resulting in a value of 1 for the lowest state resource load. A higher resource load value may suggest that resources in a given region are less likely to meet demand.

## Results

### Resource Database Validation and Categorization

We originally identified 29,935 autism resources in the United States and removed 8402 resources from the GapMap database through our validation and deduplication process. Additional resources were then extracted from Autism Speaks and Google Places, verified through our validation pipeline, merged with our database, and deduplicated, resulting in the addition of 6470 resources and amounting to a total of 28,003 unique and validated resources in the GapMap database. Table 2 shows the allocation of all resources into each of the 7 resource categories. Resources may be allocated to multiple categories.

### Distance Between Resources and Individuals

The estimated average distance between an individual with autism and the nearest autism resource belonging to any of the 7 resource categories across the United States is 17.12 km. Table 2 outlines the average distance of the nearest resource specific to each category. Due to the large number of US counties, Table 3 only shows the 10 states with the largest and smallest average distances to the nearest resource, respectively. On average across the United States, individuals are over 21 km away from resources that belong to the *Therapy* category and are farthest away from resources that belong to the *Diagnosis* category, at over 35 km. Figure 2 represents a heat map of the overall density of resources across the United States. As shown in the map, resources are sparsely distributed across the western half of the country and are densely crowded surrounding regions of higher population density.

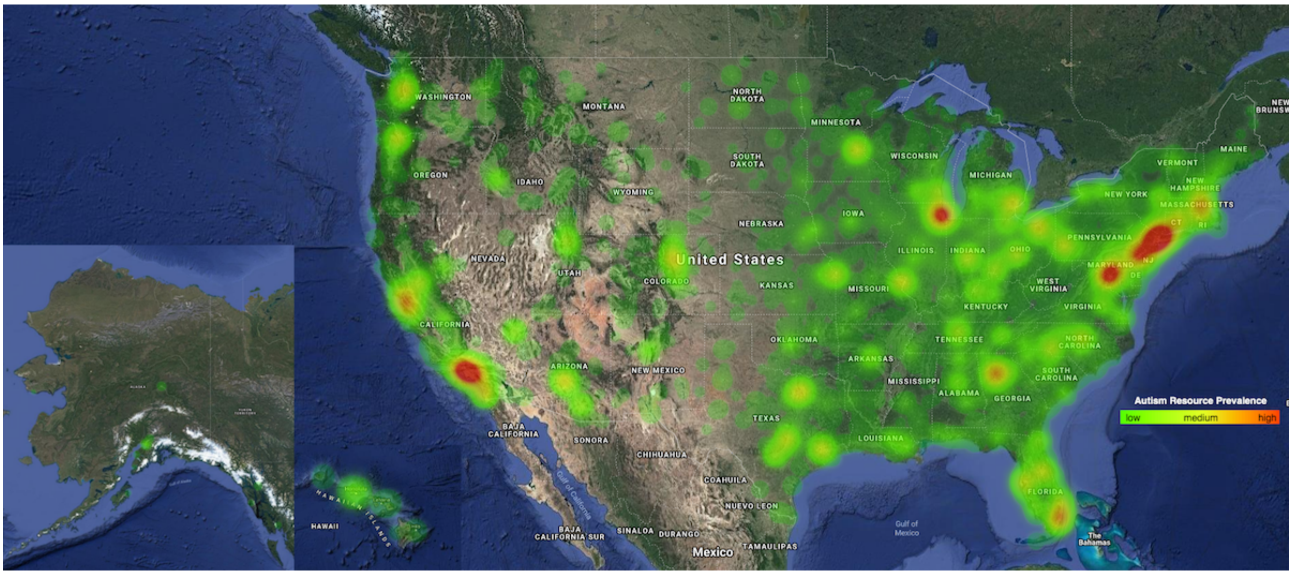
**Table 2.** The number, percent breakdown, and average distance of autism resources for each of the 7 resource categories.

| Category | Resources, n (%) | Average distance (km) |
|---|---|---|
| All resources | 28,003 (100.00) | 17.12 |
| Therapy | 11,602 (41.43) | 21.84 |
| Support | 8153 (29.11) | 21.65 |
| Health | 7236 (25.84) | 24.02 |
| Education | 5595 (19.98) | 24.17 |
| Recreation | 2791 (9.97) | 30.44 |
| Diagnosis | 2472 (8.83) | 35.49 |
| Other | 5190 (18.53) | 26.80 |

**Table 3.** Top 10 states in the United States with the largest and smallest average distances between an individual with autism and the nearest autism resource, respectively.

| Top 10 states with largest distance | | Top 10 states with smallest distance | |
|---|---|---|---|
| State | Distance (km) | State | Distance (km) |
| Alaska | 100.66 | New Jersey | 3.74 |
| Nevada | 54.16 | Connecticut | 4.67 |
| Wyoming | 54.14 | Maryland | 5.65 |
| Montana | 49.12 | Massachusetts | 6.42 |
| Arizona | 43.69 | New York | 7.07 |
| North Dakota | 41.52 | Rhode Island | 7.20 |
| New Mexico | 36.75 | Pennsylvania | 7.60 |
| South Dakota | 33.24 | Delaware | 8.09 |
| Oregon | 27.96 | Virginia | 9.06 |
| Idaho | 27.35 | New Hampshire | 9.3 |

**Figure 2.** A heat map depicting the density and distribution of autism resources across the United States (Alaska and Hawaii not to scale), where red indicates a high density of resources.




## Resource Load for Diagnostic Resources

Of the 3142 US counties surveyed, our analysis found that 2635 counties (83.86%) did not have a single diagnostic resource.

Table 4 shows the average resource load for the 10 states in the United States with the highest and lowest resource loads, respectively.

**Table 4.** Top 10 states in the United States with the highest and lowest resource load, respectively.

| Top 10 states with the highest load | | Top 10 states with the lowest load | |
| --- | --- | --- | --- |
| State | Resource load | State | Resource load |
| West Virginia | 5.71 | Montana | 1 |
| Kentucky | 4.26 | Connecticut | 1.09 |
| Maine | 4.15 | Colorado | 1.10 |
| Mississippi | 3.73 | Rhode Island | 1.10 |
| New Mexico | 3.71 | New Jersey | 1.14 |
| Oklahoma | 3.49 | Pennsylvania | 1.18 |
| South Carolina | 3.26 | Massachusetts | 1.25 |
| Nevada | 2.82 | Wisconsin | 1.26 |
| Tennessee | 2.68 | Maryland | 1.32 |
| Nebraska | 2.64 | New York | 1.32 |

## Discussion

### Principal Findings

In this study, we utilized GapMap to measure autism resource accessibility and availability across the United States. GapMap's database contains 28,003 unique, validated, and categorized autism resources. As shown in Table 2, average distance is negatively correlated with the number of resources, with the exception of *Therapy* and *Support* resources. This suggests that *Support* resources are more evenly distributed around the population than *Therapy* resources. These data show an uneven distribution between resources, which calls for the need to replace simulated population data of families affected by autism with actual population data. This will help service providers strategically allocate resources in locations, which would allow a distribution of services commensurate with the population in need.

The average distance between the nearest resource belonging to any resource category and a family affected by autism is 17.12 km. As resources that belong to an individual resource category are a subset of the total number of resources, the average distances corresponding to each individual category will be greater than 17.12 km, as the average distance is negatively correlated with the number of resources. The *Therapy* resource category contained the most resources (41.43% [11,602/28,003] of all resources on GapMap) and was tied with the *Support* resource category for the lowest average distance to families, at 21 km. The *Diagnosis* resource category contained the fewest resources (8.83% [2472/28,003] of all resources in GapMap) and had the greatest average distance (35.49 km) to individuals with autism, providing some support for the diagnostic bottleneck often reported by parents [12,14,31]. This bottleneck may be largely attributed to the lack of qualified specialists who can perform an autism diagnosis. It is estimated that there are only 8300 child psychiatrists, 1500 child neurologists, and 1000 developmental-behavioral pediatricians in the United States [32]. Specialists working with autism are even scarcer, in part, because of the lack of monetary incentives compared with other specialties; developmental-behavioral pediatricians undergo extra years of training through fellowship programs but earn equal or less than primary care physicians [14].

Of the 3142 US counties, only 507 have 1 or more diagnostic resources, suggesting that centers are not only being overloaded by its county population but also by individuals outside of the county with no access to diagnostic resources locally. This observation aligns with the historical lack of mental and behavioral health resources in rural communities [33,34]. In contrast, a state such as California no longer requires an autism diagnosis to access behavioral health treatment: families and individuals with autism may seek therapy services while waiting for a diagnosis. Further research via GapMap presents an opportunity to enhance knowledge surrounding resource use, autism prevalence, and resource availability specifically in rural areas.

In a study of several countries including the United States, Williams et al found urban locations reported higher prevalence estimates than those reported in rural and mixed locations [35], and research in Denmark showed a higher likelihood of early diagnosis and treatment in urban settings [36]. These data suggest that autism is underdiagnosed and therefore possibly treated later in rural settings. Unlike traditional epidemiological studies, GapMap is also able to accurately measure autism prevalence and resource availability across the United States in rural and urban locations, as families and service providers can directly access this database from their homes and provide the information needed. Furthermore, GapMap provides a robust, user-friendly autism resource database that depicts the real-time availability of resources for individuals with autism across the country.

### Limitations

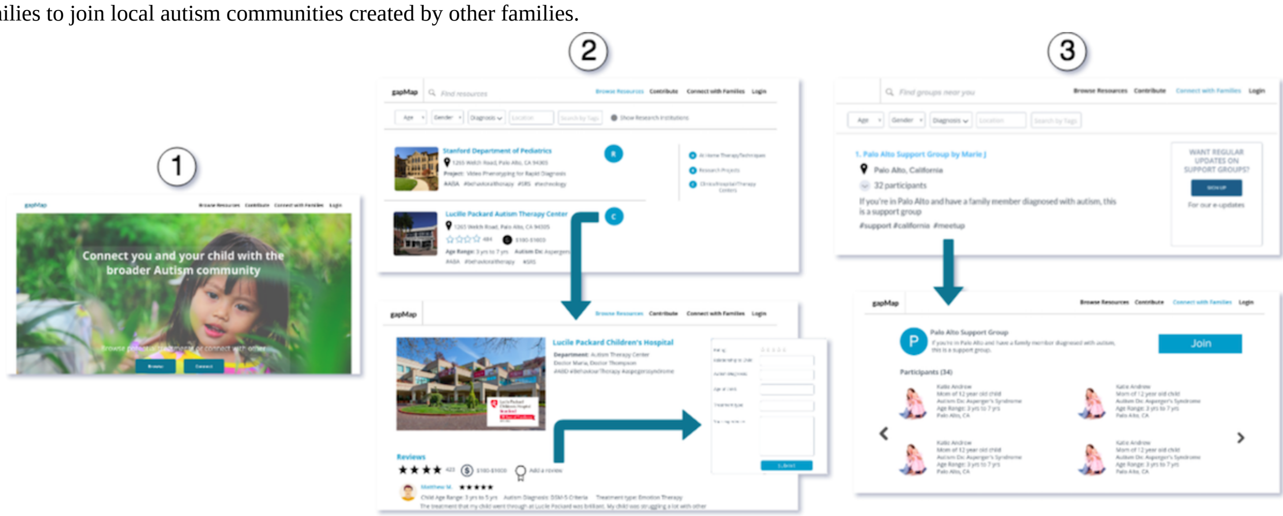
Although we made significant programmatic and manual effort to ensure the accuracy and completeness of the autism resources contained in GapMap, the true number of autism resources in the United States may be above or below the estimated value we derived through our analysis. As we do not have a method to ensure homogeneous reporting of resources by the primary databases from which we extracted our resources, the GapMap database is prone to population biases as a result of increased probability of reporting in densely crowded areas. Due to the lack of complete locational data on the population affected by autism in the United States, we assumed a randomly distributed simulation of the autism population. The geographic boundaries used to simulate the population are rectangular bounding boxes, which do not perfectly capture the actual geographic boundaries that exist in the US counties; however, we found that this primarily affects the relatively small number of counties near bodies of water. Furthermore, the resource distances are calculated with Euclidean distance, which may underestimate the true resource distance. Finally, because of a lack of annual data on the number of individuals who received an autism diagnosis and who were waitlisted for diagnosis for each state, we used a comparative resource load value, allowing only for comparison within this analysis.

## Conclusions and Future Directions

These results show an uneven distribution of autism resources throughout the United States and that diagnostic resources are the most underrepresented out of all autism resource categories analyzed. Demand for diagnostic resources varies throughout the United States, contributing to the diagnostic bottleneck, long waitlists spanning several months, and large travel distances for families in rural or underserved locations. The discrepancy in access to services across the United States could be attributed to many factors. We will attempt to parse such sources further as more granular population data are collected from families affected by autism.

Consequently, GapMap will now endeavor to collect geocoded location data and self-reported diagnoses from recruited families. Over time, this will enable a more detailed and accurate representation of gaps in access to care by enabling us to replace simulated family locations with self-reported locations. Euclidean distances between individuals and resources will be replaced with more accurate road distances that factor in street network using Google Maps API. In addition, we plan to incorporate several new features to the GapMap platform (Figure 3), which will allow families to interact with each other by rating resources, leaving reviews, adding additional resources, and submitting wait times for services. This will serve as a system that evaluates resources in terms of quality, thus providing greater incentive for families to join, while collecting real population data needed to improve autism resource epidemiological research.

**Figure 3.** Depiction of the proposed rating- and review-based features. (1) Landing page; (2) allows families to rate and review resources; (3) allows families to join local autism communities created by other families.




## Acknowledgments

The work was supported in part by funds to DPW from the National Institutes of Health (NIH; 1R01EB025025-01 & 1R21HD091500-01), The Hartwell Foundation, Bill and Melinda Gates Foundation, Coulter Foundation, Lucile Packard Foundation, and grants from Stanford's Human Centered Artificial Intelligence Program, Precision Health and Integrated Diagnostics Center (PHIND), Beckman Center, Bio-X Center, Predictives and Diagnostics Accelerator (SPADA) Spectrum, Spark Program in Translational Research, and the Wu Tsai Neurosciences Institute Neuroscience: Translate Program. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

## Conflicts of Interest

None declared.

## References

1. Hertz-Picciotto I, Delwiche L. The rise in autism and the role of age at diagnosis. Epidemiology 2009 Jan;20(1):84-90 [FREE Full text] [doi: 10.1097/EDE.0b013e3181902d15] [Medline: 19234401]

2. Christensen DL, Braun KV, Baio J, Bilder D, Charles J, Constantino JN, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - autism and developmental disabilities Monitoring network, 11 Sites, United States, 2012. MMWR Surveill Summ 2018 Nov 16;65(13):1-23 [FREE Full text] [doi: 10.15585/mmwr.ss6513a1] [Medline: 30439868]
3. Christensen DL, Bilder DA, Zahorodny W, Pettygrove S, Durkin MS, Fitzgerald RT, et al. Prevalence and characteristics of autism spectrum disorder among 4-year-old children in the autism and developmental disabilities monitoring network. J Dev Behav Pediatr 2016 Jan;37(1):1-8. [doi: 10.1097/DBP.0000000000000235] [Medline: 26651088]
4. Chiri G, Warfield ME. Unmet need and problems accessing core health care services for children with autism spectrum disorder. Matern Child Health J 2012 Jul;16(5):1081-1091. [doi: 10.1007/s10995-011-0833-6] [Medline: 21667201]
5. Kogan MD, Strickland BB, Blumberg SJ, Singh GK, Perrin JM, van Dyck PC. A national profile of the health care experiences and family impact of autism spectrum disorder among children in the United States, 2005-2006. Pediatrics 2008 Dec;122(6):e1149-e1158. [doi: 10.1542/peds.2008-1057] [Medline: 19047216]
6. Siklos S, Kerns KA. Assessing the diagnostic experiences of a small sample of parents of children with autism spectrum disorders. Res Dev Disabil 2007;28(1):9-22. [doi: 10.1016/j.ridd.2005.09.003] [Medline: 16442261]
7. US Census Bureau, Population Division. 2017. County Population Totals and Components of Change: 2010-2018 URL: https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html#par_textimage_739801612 [accessed 2018-02-28] [WebCite Cache ID 78rWBCzlM]
8. Shattuck P, Durkin M, Maenner M, Newschaffer C, Mandell DS, Wiggins L, et al. Timing of identification among children with an autism spectrum disorder: findings from a population-based surveillance study. J Am Acad Child Adolesc Psychiatry 2009 May;48(5):474-483 [FREE Full text] [doi: 10.1097/CHI.0b013e31819b3848] [Medline: 19318992]
9. Angell AM, Empey A, Zuckerman KE. Chapter four-a review of diagnosis and service disparities among children with autism from racial and ethnic minority groups in the United States. Int Rev Res Dev Disabil 2018;55:145-180. [doi: 10.1016/bs.irrdd.2018.08.003]
10. Herlihy L, Knoch K, Vibert B, Fein D. Parents' first concerns about toddlers with autism spectrum disorder: effect of sibling status. Autism 2015 Jan;19(1):20-28 [FREE Full text] [doi: 10.1177/1362361313509731] [Medline: 24216070]
11. Pierce K, Gazestani VH, Bacon E, Barnes CC, Cha D, Nalabolu S, et al. Evaluation of the Diagnostic Stability of the Early Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months. JAMA Pediatr 2019;173(6):578-587. [doi: 10.1001/jamapediatrics.2019.0624] [Medline: 31034004]
12. Mazurek MO, Handen BL, Wodka EL, Nowinski L, Butter E, Engelhardt CR. Age at first autism spectrum disorder diagnosis: the role of birth cohort, demographic factors, and clinical features. J Dev Behav Pediatr 2014;35(9):561-569. [doi: 10.1097/DBP.0000000000000097] [Medline: 25211371]
13. Dawson G, Bernier R. A quarter century of progress on the early detection and treatment of autism spectrum disorder. Dev Psychopathol 2013 Nov;25(4 Pt 2):1455-1472. [doi: 10.1017/S0954579413000710] [Medline: 24342850]
14. Gordon-Lipkin E, Foster J, Peacock G. Whittling down the wait time: exploring models to minimize the delay from initial concern to diagnosis and treatment of autism spectrum disorder. Pediatr Clin North Am 2016 Dec;63(5):851-859 [FREE Full text] [doi: 10.1016/j.pcl.2016.06.007] [Medline: 27565363]
15. Dawson G. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. Dev Psychopathol 2008;20(3):775-803. [doi: 10.1017/S0954579408000370] [Medline: 18606031]
16. Dawson G, Jones EJ, Merkle K, Venema K, Lowy R, Faja S, et al. Early behavioral intervention is associated with normalized brain activity in young children with autism. J Am Acad Child Adolesc Psychiatry 2012 Nov;51(11):1150-1159 [FREE Full text] [doi: 10.1016/j.jaac.2012.08.018] [Medline: 23101741]
17. Phillips DA, Shonkoff JP. From Neurons to Neighborhoods: The Science of Early Childhood Development. In: Washington, DC: National Academies Press. Baltimore, MD: Paul H Brookes Publishing Co; 2000.
18. Buescher A, Cidav Z, Knapp M, Mandell DS. Costs of autism spectrum disorders in the United Kingdom and the United States. JAMA Pediatr 2014 Aug;168(8):721-728. [doi: 10.1001/jamapediatrics.2014.210] [Medline: 24911948]
19. Ganz ML. The lifetime distribution of the incremental societal costs of autism. Arch Pediatr Adolesc Med 2007 Apr;161(4):343-349. [doi: 10.1001/archpedi.161.4.343] [Medline: 17404130]
20. Keenan M. The experiences of parents during diagnosis and forward planning for children with autism spectrum disorder. J Appl Res Intellect Disabil 2010;23(4):390-397. [doi: 10.1111/j.1468-3148.2010.00555.x]
21. Albert N, Daniels J, Schwartz J, Du M, Wall D. GapMap: enabling comprehensive autism resource epidemiology. JMIR Public Health Surveill 2017 May 04;3(2):e27 [FREE Full text] [doi: 10.2196/publichealth.7150] [Medline: 28473303]
22. Daniels J, Schwartz J, Albert N, Du M, Wall D. The GapMap project: a mobile surveillance system to map diagnosed autism cases and gaps in autism services globally. Mol Autism 2017;8(1):55 [FREE Full text] [doi: 10.1186/s13229-017-0163-7] [Medline: 29075431]
23. Wall DP, Du M, Ning M. GapMap. URL: https://gapmap.stanford.edu/ [accessed 2019-05-16] [WebCite Cache ID 78rOPorfz]
24. Autism Speaks. 2018. Find the Services and Support You Need URL: https://www.autismspeaks.org/directory [accessed 2018-06-15] [WebCite Cache ID 78rPYcOFS]
25. Autism Source. 2004. The Autism Society's Online Resource Database URL: http://www.autismsource.org/ [accessed 2018-03-01]

26. Parents Helping Parents. 2014. URL: https://my.php.com/resources [accessed 2018-03-01]
27. ParseHub. 2018. URL: https://www.parsehub.com/ [accessed 2018-03-01]
28. Singhal M, Shukla A. Implementation of location based services in Android using GPS and web services. International Journal of Computer Science Issues 2012 Jan;9(1):237-242 [FREE Full text]
29. TIGER/Line Shapefiles - Geography. 2017. United States Census Bureau URL: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html [accessed 2018-03-29]
30. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. MMWR Surveill Summ 2014 Mar 28;63(2):1-21 [FREE Full text] [Medline: 24670961]
31. Toomey SL, Chien AT, Elliott MN, Ratner J, Schuster MA. Disparities in unmet need for care coordination: the national survey of children's health. Pediatrics 2013 Feb;131(2):217-224. [doi: 10.1542/peds.2012-1535] [Medline: 23339228]
32. Spectrum. 2018 Nov 22. Why are there so few autism specialists? URL: https://www.spectrumnews.org/features/deep-dive/why-are-there-so-few-autism-specialists/ [accessed 2018-11-27]
33. Antezana L, Scarpa A, Valdespino A, Albright J, Richey JA. Rural trends in diagnosis and services for autism spectrum disorder. Front Psychol 2017;8:590 [FREE Full text] [doi: 10.3389/fpsyg.2017.00590] [Medline: 28473784]
34. Heflinger C, Shaw V, Higa-McMillan C, Lunn L, Brannan AM. Patterns of child mental health service delivery in a public system: rural children and the role of rural residence. J Behav Health Serv Res 2015 Jul;42(3):292-309. [doi: 10.1007/s11414-015-9464-9] [Medline: 25813915]
35. Williams JG, Higgins JP, Brayne CE. Systematic review of prevalence studies of autism spectrum disorders. Arch Dis Child 2006 Jan;91(1):8-15 [FREE Full text] [doi: 10.1136/adc.2004.062083] [Medline: 15863467]
36. Lauritsen M, Astrup A, Pedersen CB, Obel C, Schendel DE, Schieve L, et al. Urbanicity and autism spectrum disorders. J Autism Dev Disord 2014 Feb;44(2):394-404 [FREE Full text] [doi: 10.1007/s10803-013-1875-y] [Medline: 23807204]


## Abbreviations

**API:** application programming interface
**ASD:** Autism Spectrum Disorder
**AWS:** Amazon Web Services

*Edited by G Eysenbach; submitted 12.12.18; peer-reviewed by Y Cao, P Antoniou; comments to author 08.05.19; revised version received 15.05.19; accepted 16.05.19; published 10.07.19.*

*Please cite as:*
*Ning M, Daniels J, Schwartz J, Dunlap K, Washington P, Kalantarian H, Du M, Wall DP*
*Identification and Quantification of Gaps in Access to Autism Resources in the United States: An Infodemiological Study*
*J Med Internet Res 2019;21(7):e13094*
*URL: http://www.jmir.org/2019/7/e13094/*
*doi:10.2196/13094*
*PMID:31293243*

Original Paper

# Influence of Climate on Google Internet Searches for Pruritus Across 16 German Cities: Retrospective Analysis

Linda Tizek[1,2], MPH; Maximilian Schielein[1,2], MPH; Melvin Rüth[1], BSc; Sonja Ständer[3], MD; Manuel Pedro Pereira[3], MD; Bernadette Eberlein[1], MD; Tilo Biedermann[1], MD; Alexander Zink[1], MPH, MD, PhD

[1]Department of Dermatology and Allergy, School of Medicine, Technical University of Munich, Munich, Germany
[2]The Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig Maximilian University Munich, Munich, Germany
[3]Center for Chronic Pruritus, Department of Dermatology, University Hospital Münster, Muenster, Germany

**Corresponding Author:**
Alexander Zink, MPH, MD, PhD
Department of Dermatology and Allergy
School of Medicine
Technical University of Munich
Biedersteiner Str 29
Munich, 80802
Germany
Phone: 49 89 4140 3038
Fax: 49 89 4140 3552
Email: Alexander.zink@tum.de

## Abstract

**Background:** The burden of pruritus is high, especially among patients with dermatologic diseases. Identifying trends in pruritus burden and people's medical needs is challenging, since not all affected people consult a physician.

**Objective:** The purpose of this study was to investigate pruritus search behavior trends in Germany and identify associations with weather factors.

**Methods:** Google AdWords Keyword Planner was used to quantify pruritus-related search queries in 16 German cities from August 2014 to July 2018. All identified keywords were qualitatively categorized and pruritus-related terms were descriptively analyzed. The number of search queries per 100,000 inhabitants of each city was compared to environmental factors such as temperature, humidity, particulate matter 10 micrometers or less in diameter (PM10), and sunshine duration to investigate potential correlations.

**Results:** We included 1150 pruritus-related keywords, which resulted in 2,851,290 queries. "Pruritus" (n=115,680) and "anal pruritus" (n=102,390) were the most-searched-for keywords. Nearly half of all queries were related to the category *localization*, with Berlin and Munich having a comparatively high proportion of people that searched for pruritus in the genital and anal areas. People searched more frequently for information on chronic compared to acute pruritus. The most populated cities had the lowest number of queries per 100,000 inhabitants (Berlin, n=13,641; Hamburg, n=18,303; and Munich, n=21,363), while smaller cities (Kiel, n=35,027; and Freiburg, n=39,501) had the highest. Temperature had a greater effect on search query number (beta -7.94, 95% CI -10.74 to -5.15) than did PM10 (beta -5.13, 95% CI -7.04 to -3.22), humidity (beta 4.73, 95% CI 2.70 to 6.75), or sunshine duration (beta 0.66, 95% CI 0.36 to 0.97). The highest relative number of search queries occurred during the winter (ie, December to February).

**Conclusions:** By taking into account the study results, Google data analysis helps to examine people's search frequency, behavior, and interest across cities and regions. The results indicated a general increase in search queries during the winter as well as differences across cities located in the same region; for example, there was a decline in search volume in Saarbrucken, while there were increases in Cologne, Frankfurt, and Dortmund. In addition, the detected correlation between search volume and weather data seems to be valuable in predicting an increase in pruritus burden, since a significant association with rising humidity and sunshine duration, as well as declining temperature and PM10, was found. Accordingly, this is an unconventional and inexpensive method to identify search behavior trends and respective inhabitants' needs.

***(J Med Internet Res 2019;21(7):e13739)*** doi:10.2196/13739

**KEYWORDS**

pruritus; Internet; informatics; environment; weather; retrospective studies

## Introduction

Pruritus is one of the most common presenting symptoms in dermatological patients, appearing in more than 50% of patients [1,2]. In general, acute and chronic pruritus occurrences are most prevalent among people affected by scabies (up to 100%) [3], atopic dermatitis (83.3%-91.0%) [2,4,5], and psoriasis (48.6%-84.0%) [5,6]. Epidemiological studies have reported point prevalence rates in the general population ranging from 3.6% to 8.4% for acute pruritus [1,5] and 13.5% to 31.2% for chronic pruritus (ie, lasting >6 weeks) [2,7,8]. Prevalence has been observed to increase with age [9].

Pruritus can have a great impact on quality of life [10-12]. There is a higher prevalence of suicidal ideation and depression among people with severe pruritus [11,13,14]. Treatment, especially for chronic pruritus, can be very challenging owing to the diversity of underlying diseases [15,16], subjective sensations, and varying individual patient needs [17,18]. There are various scales for measuring symptom intensity and severity [19]. Pruritus is also measured using patient-oriented measurements (eg, the patient benefit index) according to the individual needs and desires of patients with respect to therapeutic outcomes [20]. A study that used the patient benefit index showed that 83.9% of patients with psoriasis considered the reduction of pruritus to be a treatment goal [21]. However, people with less severe pruritus, localized manifestations, or only occasional itching might not seek professional health care; thus, it is challenging to examine people's interest in medical needs.

An unconventional method to assess people's interest in different aspects of pruritus outside the medical setting is to analyze the volume of Internet searches for "pruritus" and accompanying expressions, since the Internet is a commonly used source of health information [22,23]. In Germany, 90% of residents use the Internet and 72% use it daily [24]. Approximately 57% of the German population have used the Internet to search for health-related information at least once [25] and the vast majority of this group (95%) use Google as their primary search engine [26]. As previously demonstrated, Google data analysis is valuable for detecting seasonal trends and making forecasts regarding various diseases, such as cancer, epilepsy, Ebola virus disease, or influenza [27-29]. For example, as previously demonstrated in the United States, search behavior was associated with incidence rates of various tumors (eg, skin cancer) [29] or coronary heart disease. In addition, analyzing Internet data is useful for examining Internet behavior, interest, and people's reactions to various incidents [27]. Studies investigating Google search volume data for the term "pruritus" in Germany have demonstrated that this method provides good insight into people's interests and needs [30,31]. However, dermatological care differs even within Germany. For example, according to the German Association of Panel Doctors (Kassenärztliche Bundesvereinigung), the supply of dermatologists (ie, the proportion of target to actual number of dermatologists as a function of regional inhabitants) was comprehensively higher in Freiburg (218.2%) and Kiel (188.1%) than in Berlin (124.1%) and Hamburg (113.3%) [32]. Besides variations in the supply of dermatologists, there are regional differences in environmental triggers, such as temperature [33-37]. Accordingly, further analysis of Google data with regard to regional differences in pruritus would be valuable. Therefore, in an effort to identify possible unmet needs, this study examined the Google search volume data of 16 German cities to investigate whether there were local differences in people's interest in pruritus and whether external factors might have had an influence on search behavior.

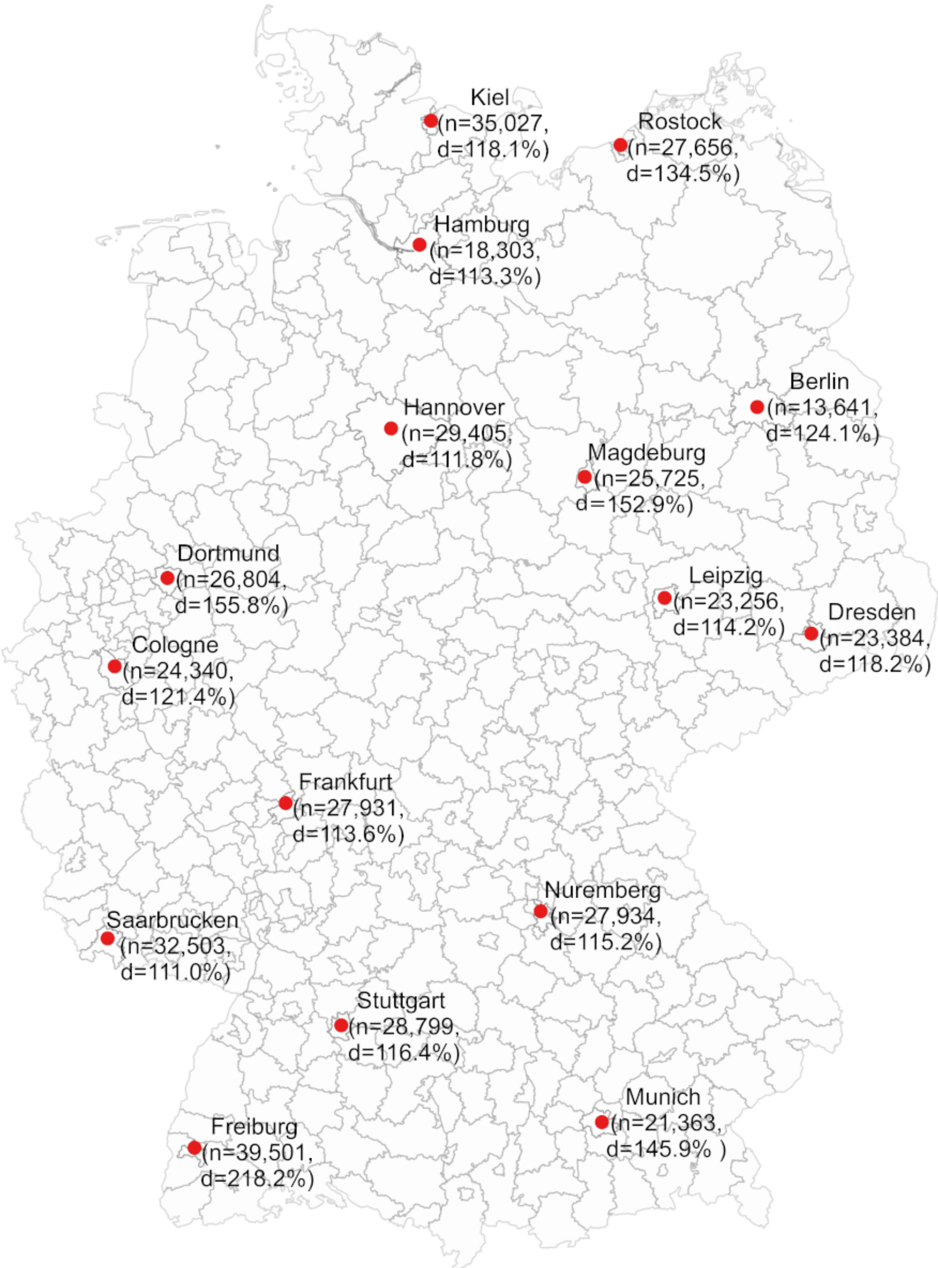
## Methods

### Study Design

A retrospective longitudinal study using Google AdWords Keyword Planner was conducted to identify the Google search volume of keywords related to pruritus in 16 cities across Germany. Four cities each from Northern Germany (Hamburg, Hannover, Kiel, and Rostock), Eastern Germany (Berlin, Leipzig, Dresden, and Magdeburg), Southern Germany (Munich, Stuttgart, Nuremberg, and Freiburg), and Western Germany (Cologne, Frankfurt, Dortmund, and Saarbrucken) were chosen for a representative evaluation to determine whether there are national and geographical differences in pruritus searches within these regions. Most of the cities were chosen because they are Germany's largest cities by population; the following 11 cities are listed in order of largest population to smallest, with their population ranking in Germany within parentheses: Berlin (1), Hamburg (2), Munich (3), Cologne (4), Frankfurt (5), Stuttgart (6), Dortmund (8), Leipzig (9), Dresden (12), Hannover (13), and Nuremberg (14). Furthermore, Kiel and Rostock were chosen to examine whether the proximity to the coast has an influence on Google search volume. The remaining cities—Freiburg, Saarbrucken, and Magdeburg—were chosen because we wanted to have a nationwide overview about various regions and they are three of the largest cities within these regions. Even though the main function of Google's Keyword Planner is to optimize advertising, it can also be used to answer scientific questions [23,30,31]. To assess search volume within a specific field, words or phrases related to the topic are initially entered into the tool. The Keyword Planner then finds the most relevant search terms. For each identified keyword, the tool provides the monthly search volume data as estimated by Google, which are available for the last 48 months. The search volume represents the total number of searches related to selected keywords [38]. In this study, the tool was used to investigate the number of search queries associated with the German lay word for pruritus or itch ("Juckreiz") from August 2014 to July 2018 (see Figure 1). The region and language settings were set so that the search volume data using Google products were limited solely to users in the abovementioned cities whose language preference was *German*.

Since the study was based on Google search terms, institutional review board approval was not needed and informed consent was not applicable.

**Figure 1.** Map showing the 16 German cities for which Google search volume analysis was performed for pruritus-related terms searched from August 2014 to July 2018. n: number of search queries per 100,000 inhabitants; d: supply of dermatologists (ie, ratio of target to actual number of dermatologists with respect to population).




## Classifications

Keywords identified by Google AdWords Keyword Planner were qualitatively analyzed. Terms that were associated with pruritus but did not explicitly include words or phrases related to pruritus (eg, "psoriasis" and "atopic dermatitis") were initially excluded from descriptive analyses. Only when investigating a correlation between the supply of general practitioners (GPs) and search volume were these terms considered, as people consulting a GP might receive other less-specific diagnoses [36]. After qualitatively assessing all pruritus-related keywords, eight categories were formed in accordance with previous studies [30,31] to determine differences in people's interests: *causes* (eg, "pruritus causes"), *conditions* (eg, "pruritus at night-time"), *influential factors* (eg, "allergic pruritus"), *localization* (eg, "itchy legs"), *pruritus descriptors* (eg, "strong pruritus"), *questions on pruritus* (eg, "what causes pruritus?"), and *treatment* (eg, "drugs against pruritus"); searches that did not fit any of these categories were placed in a *general* category (eg, "pruritus"). Terms matching several criteria were assigned to multiple categories. To assess differences in search behaviors across Germany, the search volume for each city was calculated in relation to its inhabitants [39] and then expressed as the number of search queries per 100,000 inhabitants (see Table 1).

**Table 1.** Number of searches per 100,000 inhabitants according to categories of pruritus-related keywords in 16 German cities from August 2014 to July 2018.

| City (overall search volume)[a] [39] | Number of inhabitants in 2016 | Categories and number of searches/100,000 inhabitants, n (%)[b] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Causes ($k^c$=94) | Conditions (k=148) | Influential factors (k=240) | Localization (k=499) | Pruritus descriptors (k=215) | Questions on pruritus (k=100) | Treatment (k=149) | General (k=109) |
| Berlin (N=13,641) | 3,574,830 | 779 (5.71) | 1128 (8.27) | 1966 (14.41) | 6119 (44.86) | 1537 (11.27) | 720 (5.28) | 1475 (10.82) | 2674 (19.60) |
| Hamburg (N=18,303) | 1,810,438 | 1058 (5.78) | 1665 (9.10) | 2891 (15.80) | 8083 (44.16) | 2156 (11.78) | 1012 (5.53) | 1957 (10.69) | 3262 (17.82) |
| Munich (N=21,363) | 1,464,301 | 1209 (5.66) | 1859 (8.70) | 3428 (16.04) | 9712 (45.46) | 2496 (11.68) | 1097 (5.14) | 2230 (10.44) | 3712 (17.37) |
| Cologne (N=24,340) | 1,075,935 | 1397 (5.74) | 2247 (9.23) | 3988 (16.39) | 10,889 (44.74) | 2919 (11.99) | 1357 (5.58) | 2656 (10.91) | 4040 (16.60) |
| Frankfurt (N=27,931) | 736,414 | 1589 (5.69) | 2539 (9.09) | 4645 (16.63) | 12,542 (44.90) | 3278 (11.74) | 1502 (5.38) | 3073 (11.00) | 4530 (16.22) |
| Stuttgart (N=28,799) | 628,032 | 1705 (5.92) | 2484 (8.62) | 4727 (16.42) | 13,122 (45.56) | 3439 (11.94) | 1626 (5.64) | 3064 (10.64) | 4657 (16.17) |
| Dortmund (N=26,804) | 585,813 | 1593 (5.94) | 2366 (8.83) | 4350 (16.23) | 12,069 (45.03) | 3382 (12.62) | 1726 (6.44) | 3023 (11.28) | 4198 (15.66) |
| Leipzig (N=23,256) | 571,088 | 1392 (5.99) | 1984 (8.53) | 3707 (15.94) | 10,093 (43.40) | 2884 (12.40) | 1219 (5.24) | 2590 (11.14) | 3977 (17.10) |
| Dresden (N=23,384) | 547,172 | 1482 (6.34) | 1990 (8.51) | 3540 (15.14) | 10,651 (45.55) | 2941 (12.58) | 1177 (5.03) | 2601 (11.12) | 3739 (15.99) |
| Hannover (N=29,405) | 532,864 | 1717 (5.84) | 2719 (9.25) | 4851 (16.50) | 12,941 (44.01) | 3549 (12.07) | 1687 (5.74) | 3354 (11.40) | 4643 (15.79) |
| Nuremberg (N=27,934) | 511,628 | 1667 (5.97) | 2557 (9.15) | 4535 (16.23) | 12,658 (45.31) | 3389 (12.13) | 1597 (5.72) | 3145 (11.26) | 4329 (15.50) |
| Kiel (N=35,027) | 247,441 | 2219 (6.33) | 3302 (9.43) | 5508 (15.73) | 15,749 (44.96) | 4066 (11.61) | 1908 (5.45) | 3876 (11.06) | 5428 (15.50) |
| Magdeburg (N=25,725) | 238,136 | 1655 (6.43) | 1969 (7.66) | 3523 (13.70) | 11,758 (45.71) | 3099 (12.05) | 1302 (5.06) | 3141 (12.21) | 4170 (16.21) |
| Freiburg (N=39,501) | 227,590 | 2474 (6.26) | 3203 (8.11) | 6534 (16.54) | 18,103 (45.83) | 4302 (10.89) | 2144 (5.43) | 4772 (12.08) | 5571 (14.10) |
| Rostock (N=27,656) | 207,513 | 1711 (6.19) | 2231 (8.07) | 3653 (13.21) | 12,867 (46.52) | 3301 (11.94) | 1460 (5.28) | 3446 (12.46) | 4274 (15.46) |
| Saarbrucken (N=32,503) | 179,709 | 1898 (5.84) | 2515 (7.74) | 4875 (15.00) | 15,442 (47.51) | 3734 (11.49) | 1870 (5.75) | 4018 (12.36) | 4646 (14.30) |
| Total (N=21,701) | 13,138,904 | 1272 (5.86) | 1896 (8.74) | 3409 (15.71) | 9759 (44.97) | 2566 (11.82) | 1186 (5.46) | 2392 (11.02) | 3676 (16.94) |

[a]Overall search volume/100,000 inhabitants

[b]The cumulative percentage might be over 100% since keywords could have been attributed to multiple categories.

[c]Number of keywords.

## External Factors

Data from the Climate Data Center [34] (ie, mean monthly temperature in °C, mean humidity in %, and mean monthly sunshine duration in hours) as well as atmospheric particulate matter 10 micrometers or less in diameter (PM10) data from federal states and federal government networks (Messnetze der Bundesländer sowie des Bundes) [37] were used to determine whether there were correlations with the number of search queries within each city. In addition, the correlation between searches and the supply of dermatologists and GPs [32], as well as the respective inhabitants' demographics (ie, mean age, proportion of female inhabitants, or proportion of nonnative German inhabitants) [39-41], was examined.

### Statistical Analysis

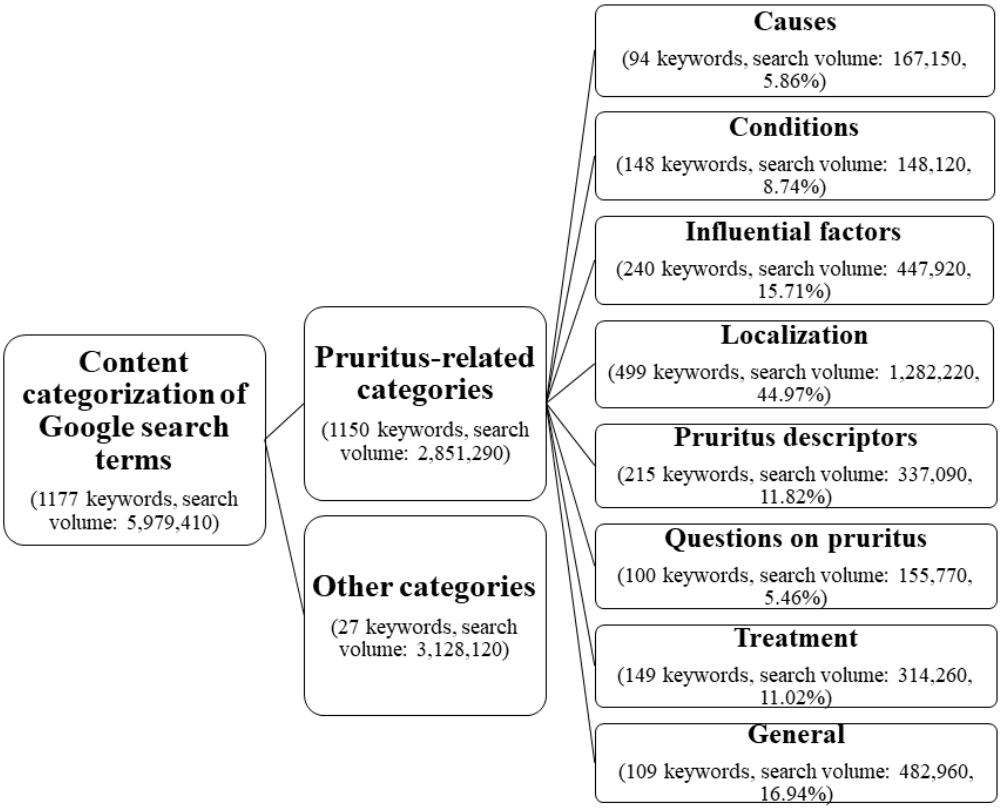
Descriptive data were generated for all categorized keywords. To determine differences in search volume per 100,000 inhabitants within cities and regions, one-way analysis of variance (ANOVA) was used. Pearson's correlation coefficient was used to assess the relationship between the number of search queries and the abovementioned external factors. Additionally, a linear regression model was generated to further assess the relationship between search queries and environmental factors. Forward selection was used to generate the best-fit model. Standardized regression coefficients (beta) and 95% CIs were estimated. Stratified analyses by region were also performed. IBM SPSS Statistics for Windows, version 25.0 (IBM Corp) was used for the statistical and spatial analyses. Geodata from the German Federal Agency for Cartography and Geodesy [42] were used to determine administrative boundaries using a geographic information system, QGIS, version 2.14.22 (QGIS Development Team).

## Results

### Overview

In total, 1177 keywords related to the German lay word for pruritus were identified in all 16 German cities. Of these, 1150 were considered for further analyses, resulting in a search volume of 2,851,290 queries. Most of the keywords were assigned to the *localization* category (499/1150, 43.39%), whereas the smallest number of keywords were categorized as *causes* (94/1150, 8.17%) (see Figure 2). The most-searched-for keywords were "pruritus" (n=115,680), "anal pruritus" (n=102,390), "pruritus on the whole body" (n=56,660), and "itchy skin" (n=53,480) (see Multimedia Appendix 1).

**Figure 2.** Content categorization of search terms identified by Google AdWords Keyword Planner.




### Comparison of 16 Cities Across Germany

As expected, the greatest number of absolute searches for pruritus-related keywords occurred in the most populated German cities: Berlin (n=487,650), Hamburg (n=331,360), and Munich (n=312,820). However, when adjusting the search volume by the number of inhabitants, Freiburg and Kiel had the highest number per 100,000 inhabitants, with 39,501 and 35,027 search queries, respectively (see Figure 1). The analyses showed that Berlin had a significantly lower number of search queries per 100,000 inhabitants (n=13,641) than all other cities except for Hamburg (n=18,303, *P*=.99) and Munich (n=21,363, *P*=.05). Overall, in Eastern Germany, the number of search queries (86,006 searches/100,000 inhabitants) was lower compared to Northern Germany (110,391 searches/100,000 inhabitants, *P*<.001), Southern Germany (117,597 searches/100,000 inhabitants, *P*<.001), and Western Germany (111,578 searches/100,000 inhabitants, *P*<.001).

About 43.4%-47.5% of all queries were related to the body parts affected by pruritus, meaning that *localization* was the category with the highest relative number of searches: 9759 searches per

100,000 inhabitants. Especially in Berlin and Munich, a high proportion of people (844/4103 [20.57%] searches/100,000 inhabitants and 1311/6519 [20.11%] searches/100,000 inhabitants, respectively) searched for pruritus in the genital or anal areas, while, overall, approximately 11.55% of *localization*-related keywords focused on these areas. A total of 2566 searches per 100,000 inhabitants included more specific *pruritus descriptors*, such as “constant” (457/2566 [17.81%] searches/100,000 inhabitants). In Berlin, the number of searches per 100,000 inhabitants for keywords including information on *pruritus descriptors* only was 1537, while more than 4000 searches were registered in Kiel and Freiburg. In general, people searched more frequently for information on chronic pruritus (80/2566 [3.12%] searches/100,000 inhabitants) compared to acute pruritus (17/2566 [0.66%] searches/100,000 inhabitants). The most-searched *influential factor* was “liver” (515/3409 [15.12%] searches/100,000 inhabitants), followed by “stress” (365/3409 [10.72%] searches/100,000 inhabitants) and “allergy” (354/3409 [10.39%] searches per 100,000 inhabitants). Of all categories, *questions on pruritus* was the category with the fewest number of searches: 1186 searches per 100,000 inhabitants. The proportion ranged from 5.0% to 6.4% among the cities, with the lowest proportion recorded in Dresden and the highest in Dortmund (see Table 1 and 2).

The vast majority of categories were negatively correlated with each other. The categories *conditions*, *influential factors*, and *general* showed a negative correlation with all other categories. Compared to that, a positive correlation was detected between the categories *questions on pruritus* and *causes* ($r=.12$, $P<.001$) as well as *treatment* ($r=.34$, $P<.001$).

## Time Course of Search Behavior

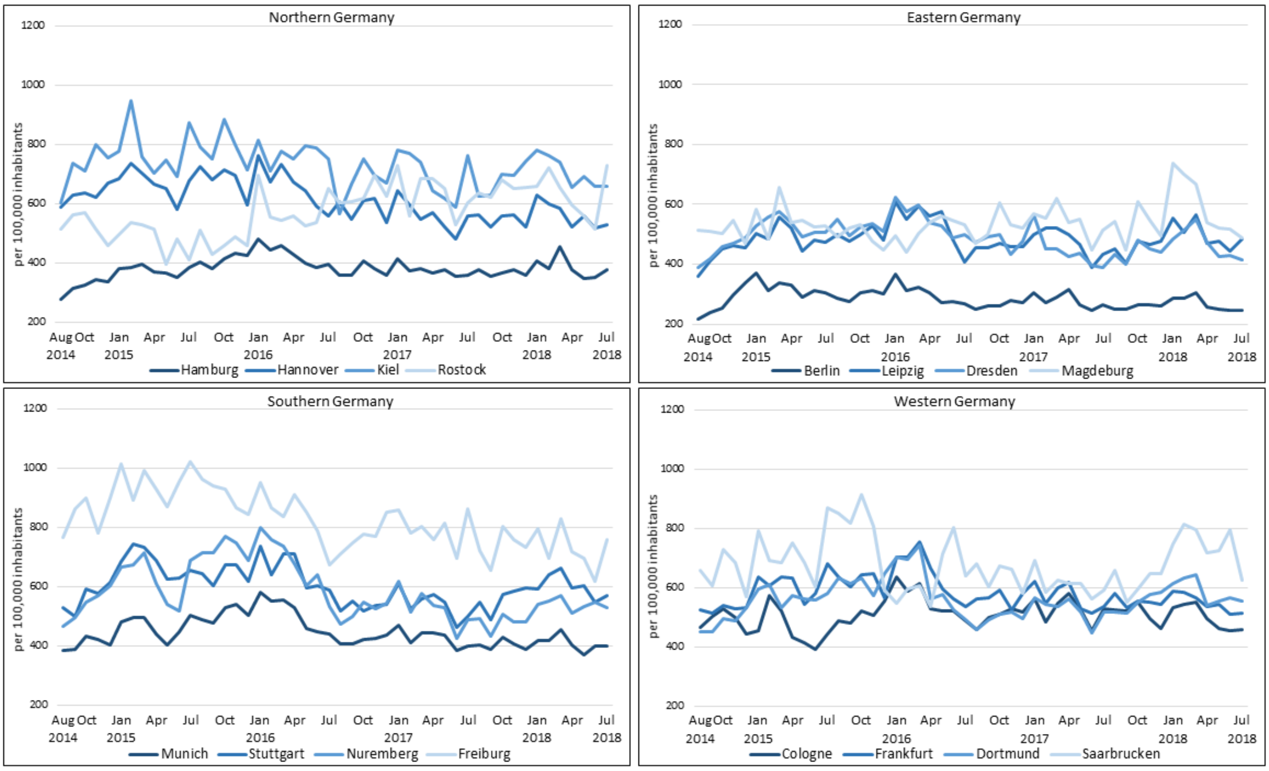
The average monthly number of searches was 452 per 100,000 inhabitants, with the greatest number occurring in Freiburg (n=823) and the smallest in Berlin (n=284). While the monthly number of searches per capita remained relatively stable in Berlin (range 239-370: 131) and Hamburg (range 316-482: 166), high seasonal ranges were observed in Kiel (range 566-946: 390) and Freiburg (range 620-1019: 399). Interestingly, a major decrease in search queries was observed in Saarbrucken from October 2015 to April 2016, whereas it was increased in all other cities in Western Germany. Except for Frankfurt, Dortmund, Freiburg, and Saarbrucken, the vast majority of cities had the highest relative number of search queries during the winter (ie, December to February). Across the four aberrant cities, the greatest numbers of searches were observed in March 2016 in Frankfurt (n=754), March 2016 in Dortmund (n=743), July 2015 in Freiburg (n=1019), and October 2015 in Saarbrucken (n=913). Overall, the highest number of relative searches occurred in January 2016 (n=562) and the lowest occurred in August 2014 (n=378) (see Figure 3).

**Table 2.** The five most-searched-for terms within each category across all examined cities, expressed as search queries per 100,000 inhabitants.

| Category and search terms | n[a] (%) |
| --- | --- |
| **Conditions (N=1896)** | |
| Most common: During night-time | 666 (35.12) |
| Second-most common: During the evening | 499 (26.30) |
| Third-most common: During pregnancy | 217 (11.46) |
| Fourth-most common: Heat-induced | 94 (4.97) |
| Fifth-most common: During winter | 87 (4.60) |
| **Influencing factors (N=3409)** | |
| Most common: Liver | 515 (15.12) |
| Second-most common: Stress | 365 (10.72) |
| Third-most common: Allergies | 354 (10.39) |
| Fourth-most common: Diabetes | 331 (9.70) |
| Fifth-most common: Dry skin | 319 (9.34) |
| **Localization (N=9759)** | |
| Most common: Whole body | 3375 (34.59) |
| Second-most common: Legs | 1516 (15.53) |
| Third-most common: Anal or genital area | 1127 (11.55) |
| Fourth-most common: Hands | 877 (8.98) |
| Fifth-most common: Head | 791 (8.11) |
| **Pruritus descriptors (N=2566)** | |
| Most common: Strong | 812 (31.65) |
| Second-most common: Constant | 457 (17.81) |
| Third-most common: Sudden | 392 (15.26) |
| Fourth-most common: Extreme | 286 (11.16) |
| Fifth-most common: Burning | 142 (5.54) |
| **Questions on pruritus (N=1186)** | |
| Most common: What helps against pruritus? | 120 (10.12) |
| Second-most common: What helps against pruritus on the whole body? | 45 (3.79) |
| Third-most common: Why does the skin develop pruritus? | 45 (3.79) |
| Fourth-most common: What helps against itchy skin? | 41 (3.45) |
| Fifth-most common: How does pruritus occur? | 38 (3.20) |
| **Treatment (N=2392)** | |
| Most common: General | 1205 (50.36) |
| Second-most common: Ointment | 290 (12.14) |
| Third-most common: Cream | 271 (11.31) |
| Fourth-most common: Drugs | 204 (8.54) |
| Fifth-most common: Home remedies | 179 (7.48) |

[a]Number of searches/100,000 inhabitants.

**Figure 3.** Trends in Google search volume per 100,000 inhabitants for pruritus-related keywords in 16 German cities from August 2014 to July 2018.




## Supply of Dermatologists and General Practitioners

A high correlation between the number of search queries per 100,000 inhabitants and the supply of dermatologists (r=.56, *P*=.02) was identified (see Figure 1). This correlation was stronger than the correlation with the supply of GPs (r=.34, *P*=.19). When search terms that did not explicitly include pruritus (eg, “psoriasis” or “atopic eczema”) were also included in the analyses, the correlation was .51 (*P*=.05) for the supply of dermatologists and .50 (*P*=.05) for that of GPs.

## Inhabitants’ Demographics

The number of search queries per 100,000 inhabitants was negatively correlated with the respective inhabitants’ mean age (r=-.12, *P*=.67) and the proportion of nonnative Germans (r=-.05, *P*=.85). In contrast, the proportion of female searchers showed a positive association (r=.33, *P*=.21). However, none of these correlations were significant.

## Environmental Factors

Overall, a negative correlation was observed between searches and temperature (r=-.14, *P*<.001). This correlation was especially high in Berlin and Leipzig (r=-.59, *P*<.001 and r=-.52, *P*<.001, respectively). In contrast to temperature, monthly relative humidity was positively correlated with the searches, indicating that the number of search queries increased with higher humidity (r=.15, *P*<.001). The most significant correlations with humidity were detected in Berlin (r=.32, *P*=.03), Cologne (r=.33, *P*=.02), and Kiel (r=.35, *P*=.02) (see Table 3).

Overall, the monthly mean daily temperature had the greatest effect on the number of search queries per 100,000 inhabitants, with lower temperatures resulting in a higher number of search queries (beta -7.94, 95% CI -10.74 to -5.15). This effect was particularly high in Eastern Germany (beta -11.74, 95% CI -15.94 to -7.54) and Western Germany (beta -9.23, 95% CI -12.5 to 5.92). The level of PM10 was also associated with a highly negative effect (beta -5.13, 95% CI -7.04 to -3.22), whereas relative humidity (beta 4.72, 95% CI 2.70 to 6.75) and the duration of sunshine (beta 0.66, 95% CI 0.36 to 0.97) showed a positive effect. However, when analyzing these influences with respect to different regions, PM10 was found to exert a highly positive effect in Northern Germany (beta 6.42, 95% CI 2.41 to 10.43). In contrast, a negative effect was observed in all the other regions: Eastern Germany (beta -7.33, 95% CI -10.42 to -4.25), Southern Germany (beta -3.50, 95% CI -6.65 to -0.35), and Western Germany (beta -8.14, 95% CI -10.75 to -5.53) (see Table 4).

**Table 3.** Correlation between the number of search queries and selected environmental factors across 16 German cities from August 2014 to July 2018.

| Cities | Monthly temperature (°C) | | | Monthly humidity (%) | | | Monthly sunshine duration (hours) | | | Monthly PM10[a] (μg/m$^3$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean (SD) | r[b] | P | Mean (SD) | r | P | Mean (SD) | r | P | Mean (SD) | r | P |
| Berlin | 11.0 (7.0) | -.59 | <.001 | 73.0 (9.9) | .32 | .03 | 140.1 (81.0) | -.41 | .003 | 23.9 (5.9) | .28 | .06 |
| Hamburg | 10.1 (9.2) | -.44 | .002 | 80.4 (7.1) | .18 | .22 | 131.5 (76.8) | -.33 | .02 | 20.0 (4.5) | .11 | .48 |
| Munich | 10.7 (7.1) | -.33 | .02 | 72.8 (7.5) | .07 | .63 | 146.9 (74.0) | -.20 | .18 | 20.5 (6.8) | .27 | .06 |
| Cologne | 12.0 (6.1) | -.47 | .001 | 74.9 (8.3) | .33 | .02 | 125.2 (66.2) | -.41 | .004 | 23.3 (4.1) | .13 | .37 |
| Frankfurt | 11.4 (6.8) | -.36 | .01 | 74.2 (10.1) | .04 | .80 | 133.3 (77.6) | -.18 | .24 | 21.3 (5.1) | .06 | .70 |
| Stuttgart | 11.3 (6.8) | -.41 | .003 | 72.6 (8.3) | .01 | .97 | 148.0 (75.3) | -.29 | .046 | 19.9 (6.7) | .12 | .40 |
| Dortmund | 11.0 (6.8) | -.47 | .001 | 77.1 (6.4) | .10 | .50 | 128.6 (63.2) | -.28 | .06 | 23.0 (4.9) | .004 | .98 |
| Leipzig | 10.6 (6.8) | -.52 | <.001 | 75.9 (8.0) | .09 | .56 | 137.8 (72.9) | -.29 | .049 | 21.0 (5.8) | .38 | .007 |
| Dresden | 11.1 (6.9) | -.48 | .001 | 73.2 (7.8) | .23 | .12 | 136.1 (69.8) | -.24 | .10 | 21.1 (5.7) | .50 | <.001 |
| Hannover | 10.5 (6.2) | -.38 | .008 | 78.4 (7.5) | .21 | .15 | 127.2 (72.3) | -.27 | .07 | 19.1 (4.2) | .35 | .01 |
| Nuremberg | 10.4 (7.0) | -.31 | .03 | 75.9 (9.6) | .12 | .42 | 144.1 (81.1) | -.32 | .03 | 24.0 (8.4) | .33 | .02 |
| Kiel | 10.0 (5.9) | -.38 | .009 | 82.5 (5.7) | .35 | .02 | 135.0 (90.7) | -.37 | .009 | 21.9 (5.8) | .54 | <.001 |
| Magdeburg | 10.8 (6.8) | -.33 | .02 | 75.9 (8.7) | .11 | .45 | 142.5 (78.4) | -.14 | .36 | 20.3 (4.8) | .04 | .80 |
| Freiburg | 11.4 (6.7) | -.19 | .20 | 75.9 (7.3) | .06 | .67 | 149.9 (76.3) | -.14 | .36 | 16.0 (4.8) | .30 | .04 |
| Rostock | 10.2 (6.1) | -.22 | .13 | 79.0 (5.2) | .22 | .13 | 150.8 (90.6) | -.18 | .21 | 19.8 (5.1) | .04 | .79 |
| Saarbrucken | 11.2 (6.5) | .08 | .57 | 78.2 (9.3) | -.22 | .14 | 134.4 (80.7) | .04 | .80 | 18.0 (4.2) | -.05 | .76 |
| All cities | 10.9 (6.5) | -.14 | <.001 | 76.2 (8.4) | .15 | <.001 | 138.2 (76.7) | -.09 | .01 | 20.8 (5.9) | -.08 | .04 |

[a]PM10: particulate matter 10 micrometers or less in diameter.

[b]Pearson's correlation coefficient (r) was used to assess the correlation between the number of search queries and selected environmental factors within the city.

**Table 4.** Results of the linear regression using forward selection to assess the relationship between environmental factors and number of search queries per 100,000 inhabitants.

| Covariates | Linear regression results for each German region, OR[a] (95% CI) | | | | |
|---|---|---|---|---|---|
| | Overall | Northern Germany | Eastern Germany | Southern Germany | Western Germany |
| PM10[b] | -5.13 (-7.04 to -3.22) | 6.42 (2.41 to 10.43) | -7.33 (-10.42 to -4.25) | -3.50 (-6.65 to -0.35) | -8.14 (-10.75 to -5.54) |
| Temperature | -7.94 (-10.74 to -5.15) | N/A[c] | -11.74 (-15.94 to -7.54) | N/A | -9.23 (-12.5 to 5.92) |
| Sunshine duration | 0.66 (0.36 to 0.97) | N/A | 1.25 (0.71 to 1.79) | N/A | 0.40 (0.13 to 0.67) |
| Humidity | 4.72 (2.70 to 6.75) | N/A | 7.07 (3.41 to 10.72) | 3.58 (0.80 to 6.36) | N/A |

[a]OR: odds ratio.

[b]PM10: particulate matter 10 micrometers or less in diameter.

[c]Not applicable; since a linear regression using forward selection was performed, some of the variables were not significant within each region.

## Discussion

### Principal Findings

From August 2014 to July 2018, 2,851,290 Google searches were performed for pruritus in the 16 examined German cities. Overall, the most common search terms were "pruritus," "anal pruritus," "pruritus on the whole body," and "itchy skin." The analyses showed that Berlin, Hamburg, and Munich, Germany's three most-populated cities, had a comprehensively lower number of search queries per 100,000 inhabitants compared with smaller cities like Saarbrucken, Kiel, and Freiburg. In almost all cities, the highest number of searches were observed in the winter, with most occurring in January 2016 and the least number of searches occurring in August 2014.

There is some evidence that Google search analyses are an effective and inexpensive tool for assessing disease trends, such as seasonal variation in multiple medical fields [27,29,43]. Previous studies using Google Trends data demonstrated correlations with various sources of data, such as coronary heart disease epidemiology [44], cancer incidence and mortality rates [29], and cases of Ebola virus disease [45]. Each of the studies showed a positive correlation, which suggested that this is an alternative solution for disease surveillance [43]. This study investigated a possible correlation between search volume and the supply of physicians. The results indicated that the number of search queries showed a higher correlation with the supply of dermatologists (r=.56) than with the supply of GPs (r=.34). This was also the case when considering search terms, such as "psoriasis" or "atopic eczema." One explanation for this discrepancy could be that, comparable to the diagnosis of xerosis [36], people consulting a dermatologist might receive a specific diagnosis for their pruritus and, therefore, may have performed explicit searches for pruritus. However, people seeing GPs might receive other less-specific diagnoses. Otherwise, this could indicate that, in spite of a high supply of physicians, there is still a definite need for medical care.

This study also revealed that the number of searches was correlated with climate. Interestingly, the effect of weather data differed within the regions. For example, in Northern Germany, only PM10 was found to have a significant effect on search behavior, while in Eastern Germany, all environmental factors examined were significant. Additionally, PM10 had a positive effect on the number of search queries in Northern Germany, which was contrary to all other regions. Although this study points to a negative effect from PM10, a positive effect, as it was observed in Northern Germany, is feasible since previous studies reported an association with pruritus or eczema [46]. Humidity and duration of sunshine were found to have a consistent positive effect across all regions. Temperature, in particular, was found to have a great effect on the number of search queries. It was found that a 1°C decrease in average monthly temperature correlated with around 8 more search queries per 100,000 inhabitants, with a greater effect observed in Eastern and Western Germany. Previous studies already reported that temperature influences the occurrence of pruritus [33,35]; for example, pruritus caused by xerosis is more common in colder months [36,47]. Similar to other German studies [30,31], we found an overall negative correlation between temperature and search volume across all cities during the entire study period (r=-.14). A study in the United States and the United Kingdom, however, reported contrary results ($r_s$=.42 and $r_s$=.27, respectively) [48]. Since no evidence was found that the United States or the United Kingdom generally have a warmer climate, search behavior has to be influenced by several other factors, for example, populations' demographics, health campaigns, or medial visibility of diseases [23,49,50]. Despite an overall negative correlation with temperature, there were a few cities with a strikingly high number of searches in warmer months: Kiel, Freiburg, and Saarbrucken in July 2015. However, no considerable search-related differences according to temperature, humidity, PM10, or cumulative sunshine duration were found in comparison with other years. Thus, it confirms that there must be additional variables (eg, prevalence of allergies and pollen season) that influence search behaviors. This seems to be supported by the fact that a strong decline in search volume was observed in Saarbrucken from October 2015 to April 2016, while there were increases in Cologne, Dortmund, and Frankfurt, even though no significant differences in environmental factors were identified.

Besides examining disease trends, Google data can also be used for gaining insights into people's health information-seeking behaviors [23,30,31]. Previous studies examining Google search volumes of pruritus-related keywords across Germany showed that the vast majority of searches (72.0%-72.6%) focused on *influential factors* [30,31], whereas most of the searches we found were related to *localization* (43.4%-47.5%) and only a small proportion were focused on *influential factors* (13.2%-16.6%). These differences might have occurred because this study only considered urban populations and search terms that explicitly included words or phrases related to pruritus, while other studies considered a larger population and additional keywords such as "psoriasis" and "atopic eczema." Apart from this discrepancy, our results are mainly consistent with others in the literature. A differentiated consideration of the cities further enables the identification of general differences in medical needs or for a specific condition. When setting the search volume in relation to the respective inhabitants, we found that smaller cities were more likely to have a higher number of search queries. The highest relative search volume was observed in Freiburg, which might be somewhat influenced by the highest proportion of female inhabitants and lowest mean age in comparison to all other cities [39,41]. The results also indicated that people living in Berlin searched less frequently for information, and those who used Google were more likely to use unspecific keywords, which is expressed by the highest proportion of general keywords. At the same time, people living in Berlin or Munich seemed to have an especially high interest in anal pruritus, as 20.57% and 20.11%, respectively, of all *localization*-related searches focused on anal pruritus (overall 11.55%). If prevention campaigns could be adapted to regional differences, they might help to better address people's therapeutic needs.

## Limitations

There are some limitations to this study that should be discussed. Google analyses are restricted to people who have access to the Internet and use Google as a search engine. In general, younger people use the Internet more frequently [24]. Since the Google AdWords Keyword Planner does not provide information on users' general demographics, no statement about the examined population can be made. Although no clear correlation between the inhabitants' mean age, percentage of female inhabitants, or percentage of nonnative Germans was found in this study, the study results might be somewhat influenced by them. On the one hand, this might have led to an underestimation of the pruritus burden, which is actually more prevalent among older people [9]. On the other hand, differences in the number of searches could result from the age distribution of Google users across the cities. Moreover, the search volume could have been influenced by different proportions of nonnative German speakers within the cities, as the study setting was limited to people whose language of preference was German, with respect to Google use and German search terms. A further limitation was that Google provides automatic completion of search terms, which might bias people's search behavior.

## Conclusions

By taking into account the study objective and results, analyses of Google data are extremely useful for medical care since the public's search behavior, including interest and frequency, could be investigated. Moreover, the study found that, for example, the number of search queries was negatively correlated with temperature and PM10, whereas it was positively correlated with humidity and sunshine duration. Considering this information, the analyses seemed to be valuable in forecasting higher public interest and need when weather is changing. In addition to this study, future studies could focus on age- or sex-specific aspects of dermatologic conditions to better target population-specific health care needs by implementing specifically adapted public health campaigns.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

List of keywords and number of times searched.

[PDF File (Adobe PDF File), 45KB - jmir_v21i7e13739_app1.pdf ]

## References


1. Weisshaar E. Epidemiology of itch. Curr Probl Dermatol 2016;50:5-10. [doi: 10.1159/000446010] [Medline: 27578064]
2. Schut C, Dalgard FJ, Halvorsen JA, Gieler U, Lien L, Aragones LT, et al. Occurrence, chronicity and intensity of itch in a clinical consecutive sample of patients with skin diseases: A multi-centre study in 13 European countries. Acta Derm Venereol 2019 Feb 01;99(2):146-151 [FREE Full text] [doi: 10.2340/00015555-3040] [Medline: 30226526]
3. Stamm LV, Strowd LC. Ignoring the "itch": The global health problem of scabies. Am J Trop Med Hyg 2017 Dec;97(6):1647-1649 [FREE Full text] [doi: 10.4269/ajtmh.17-0242] [Medline: 29016295]
4. Dawn A, Papoiu AD, Chan YH, Rapp SR, Rassette N, Yosipovitch G. Itch characteristics in atopic dermatitis: Results of a Web-based questionnaire. Br J Dermatol 2009 Mar;160(3):642-644. [doi: 10.1111/j.1365-2133.2008.08941.x] [Medline: 19067703]
5. Kopyciok ME, Ständer HF, Osada N, Steinke S, Ständer S. Prevalence and characteristics of pruritus: A one-week cross-sectional study in a German dermatology practice. Acta Derm Venereol 2016 Jan;96(1):50-55 [FREE Full text] [doi: 10.2340/00015555-2166] [Medline: 26067841]
6. Yosipovitch G, Goon A, Wee J, Chan YH, Goh CL. The prevalence and clinical characteristics of pruritus among patients with extensive psoriasis. Br J Dermatol 2000 Nov;143(5):969-973. [Medline: 11069504]
7. Matterne U, Apfelbacher CJ, Loerbroks A, Schwarzer T, Büttner M, Ofenloch R, et al. Prevalence, correlates and characteristics of chronic pruritus: A population-based cross-sectional study. Acta Derm Venereol 2011 Oct;91(6):674-679 [FREE Full text] [doi: 10.2340/00015555-1159] [Medline: 21879245]
8. Matterne U, Apfelbacher CJ, Vogelgsang L, Loerbroks A, Weisshaar E. Incidence and determinants of chronic pruritus: A population-based cohort study. Acta Derm Venereol 2013 Sep 04;93(5):532-537 [FREE Full text] [doi: 10.2340/00015555-1572] [Medline: 23450324]
9. Ständer S, Schäfer I, Phan NQ, Blome C, Herberger K, Heigel H, et al. Prevalence of chronic pruritus in Germany: Results of a cross-sectional study in a sample working population of 11,730. Dermatology 2010;221(3):229-235. [doi: 10.1159/000319862] [Medline: 20924157]
10. Marron SE, Tomas-Aragones L, Boira S, Campos-Rodenas R. Quality of life, emotional well-being and family repercussions in dermatological patients experiencing chronic itching: A pilot study. Acta Derm Venereol 2016 Mar;96(3):331-335 [FREE Full text] [doi: 10.2340/00015555-2263] [Medline: 26524077]
11. Pereira MP, Ständer S. Assessment of severity and burden of pruritus. Allergol Int 2017 Jan;66(1):3-7 [FREE Full text] [doi: 10.1016/j.alit.2016.08.009] [Medline: 27634668]
12. Ständer S, Legat F. Pruritus: Future research is to scratch before it starts itching [Article in German]. Hautarzt 2018 Aug;69(8):618-619. [doi: 10.1007/s00105-018-4225-3] [Medline: 30083775]
13. Halvorsen JA, Dalgard F, Thoresen M, Bjertness E, Lien L. Itch and pain in adolescents are associated with suicidal ideation: A population-based cross-sectional study. Acta Derm Venereol 2012 Sep;92(5):543-546 [FREE Full text] [doi: 10.2340/00015555-1251] [Medline: 22068440]
14. Balieva F, Kupfer J, Lien L, Gieler U, Finlay AY, Tomás-Aragonés L, et al. The burden of common skin diseases assessed with the EQ5D™: A European multicentre study in 13 countries. Br J Dermatol 2017 May;176(5):1170-1178. [doi: 10.1111/bjd.15280] [Medline: 28032340]
15. Weisshaar E, Szepietowski JC, Darsow U, Misery L, Wallengren J, Mettang T, et al. European guideline on chronic pruritus. Acta Derm Venereol 2012 Sep;92(5):563-581 [FREE Full text] [doi: 10.2340/00015555-1400] [Medline: 22790094]

16. Ständer S, Weisshaar E, Mettang T, Szepietowski JC, Carstens E, Ikoma A, et al. Clinical classification of itch: A position paper of the International Forum for the Study of Itch. Acta Derm Venereol 2007;87(4):291-294 [FREE Full text] [doi: 10.2340/00015555-0305] [Medline: 17598029]
17. Steinke S, Bruland P, Blome C, Osada N, Dugas M, Fritz F, et al. Chronic pruritus: Evaluation of patient needs and treatment goals with a special regard to differences according to pruritus classification and sex. Br J Dermatol 2017 Feb;176(2):363-370. [doi: 10.1111/bjd.14910] [Medline: 27479717]
18. Dalgard F, Holm J, Svensson A, Kumar B, Sundby J. Self-reported skin morbidity and ethnicity: A population-based study in a western community. BMC Dermatol 2007 Jun 29;7:4 [FREE Full text] [doi: 10.1186/1471-5945-7-4] [Medline: 17603893]
19. Ständer S, Blome C, Anastasiadou Z, Zeidler C, Jung KA, Tsianakas A, et al. Dynamic pruritus score: Evaluation of the validity and reliability of a new instrument to assess the course of pruritus. Acta Derm Venereol 2017 Feb 08;97(2):230-234 [FREE Full text] [doi: 10.2340/00015555-2494] [Medline: 27349416]
20. Augustin M, Radtke MA, Zschocke I, Blome C, Behechtnejad J, Schäfer I, et al. The patient benefit index: A novel approach in patient-defined outcomes measurement for skin diseases. Arch Dermatol Res 2009 Sep;301(8):561-571. [doi: 10.1007/s00403-009-0928-8] [Medline: 19326133]
21. Blome C, Gosau R, Radtke MA, Reich K, Rustenbach SJ, Spehr C, et al. Patient-relevant treatment goals in psoriasis. Arch Dermatol Res 2016 Mar;308(2):69-78. [doi: 10.1007/s00403-015-1613-8] [Medline: 26688112]
22. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: Results from the US National Health Interview Survey. J Med Internet Res 2015 Apr 29;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]
23. Seidl S, Schuster B, Rüth M, Biedermann T, Zink A. What do Germans want to know about skin cancer? A nationwide Google search analysis from 2013 to 2017. J Med Internet Res 2018 Dec 02;20(5):e10327 [FREE Full text] [doi: 10.2196/10327] [Medline: 29698213]
24. Koch W, Frees B. Media Perspektiven. 2017 Sep. ARD/ZDF-Onlinestudie: Neun von zehn Deutschen online URL: http://www.ard-zdf-onlinestudie.de/files/2017/Artikel/917_Koch_Frees.pdf [accessed 2018-07-05]
25. Directorate-General for Communications Networks, Content and Technology (European Commission). European Citizens' Digital Health Literacy. Brussels, Belgium: European Union; 2015 Jan 22. URL: https://publications.europa.eu/en/publication-detail/-/publication/fd42f9e7-937c-41f3-bf03-4221b2db712b [accessed 2018-08-02]
26. Statista. Statistiken zu suchmaschinen (Search engine statistics) URL: https://de.statista.com/themen/111/suchmaschinen/ [accessed 2018-07-06]
27. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: Systematic review. J Med Internet Res 2018 Dec 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]
28. 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]
29. Wehner MR, Nead KT, Linos E. Correlation among cancer incidence and mortality rates and Internet searches in the United States. JAMA Dermatol 2017 Dec 01;153(9):911-914 [FREE Full text] [doi: 10.1001/jamadermatol.2017.1870] [Medline: 28658470]
30. Zink A, Rüth M, Schuster B, Darsow U, Biedermann T, Ständer S. Pruritus in Germany: A Google search engine analysis [Article in German]. Hautarzt 2019 Jan;70(1):21-28. [doi: 10.1007/s00105-018-4215-5] [Medline: 29876613]
31. Zink A, Schuster B, Rüth M, Pereira MP, Philipp-Dormston WG, Biedermann T, et al. Medical needs and major complaints related to pruritus in Germany: A 4-year retrospective analysis using Google AdWords Keyword Planner. J Eur Acad Dermatol Venereol 2019 Jan;33(1):151-156. [doi: 10.1111/jdv.15200] [Medline: 30067284]
32. Gesundheitsdaten. Berlin, Germany: Kassenärztliche Bundesvereinigung; 2017. Versorgungsgrade in den planungsbereichen URL: http://gesundheitsdaten.kbv.de/cms/html/17016.php [accessed 2018-10-12]
33. Lewis Z, George DN, Cowdell F, Holle H. Effects of short-term temperature change in the innocuous range on histaminergic and non-histaminergic acute itch. Acta Derm Venereol 2019 Feb 01;99(2):188-195 [FREE Full text] [doi: 10.2340/00015555-3077] [Medline: 30358890]
34. Deutscher Wetterdienst. Offenbach, Germany: Deutscher Wetterdienst Climate Data Center URL: https://cdc.dwd.de/portal/201904171234/index.html [accessed 2018-12-04]
35. Vocks E, Busch R, Fröhlich C, Borelli S, Mayer H, Ring J. Influence of weather and climate on subjective symptom intensity in atopic eczema. Int J Biometeorol 2001 Feb;45(1):27-33. [Medline: 11411412]
36. Augustin M, Wilsmann-Theis D, Körber A, Kerscher M, Itschert G, Dippel M, et al. Positionspapier: Diagnostik und therapie der xerosis cutis. J Dtsch Dermatol Ges 2018 Jul;16 Suppl 4:3-35. [doi: 10.1111/ddg.13580] [Medline: 29989340]
37. Umwelt Bundesamt. Current concentrations of air pollutants in Germany URL: https://www.umweltbundesamt.de/en/data/current-concentrations-of-air-pollutants-in-germany#/start?s=q64FAA==&_k=p5zoik [accessed 2018-10-11]
38. Google Ads. Keyword Planner: Reach the right customers with the right keywords URL: https://ads.google.com/intl/en/home/tools/keyword-planner/ [accessed 2018-07-05]

39. Statista. Die größten städte in Deutschland nach einwohnerzahl zum 31. Dezember 2017 URL: https://de.statista.com/statistik/daten/studie/1353/umfrage/einwohnerzahlen-der-grossstaedte-deutschlands/ [accessed 2018-07-20]
40. Statistisches Bundesamt (Destatis). 2018. Migration. Integration. Regionen: Ausländeranteil URL: https://service.destatis.de/DE/karten/migration_integration_regionen.html [accessed 2018-12-13]
41. Federal Institute for Research on Building, Urban Affairs and Spatial Development. Germany is growing – But not everywhere URL: https://www.bbsr.bund.de/BBSR/EN/Home/top-news/2018-germany-is-growing.html [accessed 2019-05-14]
42. Bundesamt für Kartographie und Geodäsie. Open Data - Freie daten und dienste des BKG URL: https://tinyurl.com/y8kwsu4d [accessed 2018-04-04]
43. Ling R, Lee J. Disease monitoring and health campaign evaluation using Google search activities for HIV and AIDS, stroke, colorectal cancer, and marijuana use in Canada: A retrospective observational study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]
44. Senecal C, Widmer R, Lerman L, Lerman A. Association of search engine queries for chest pain with coronary heart disease epidemiology. JAMA Cardiol 2018;3(12):1218-1221. [doi: 10.1001/jamacardio.2018.3459] [Medline: 30422176]
45. Alicino C, Bragazzi NL, Faccio V, Amicizia D, Panatto D, Gasparini R, et al. Assessing Ebola-related Web search behaviour: Insights and implications from an analytical study of Google Trends-based query volumes. Infect Dis Poverty 2015 Dec 10;4:54 [FREE Full text] [doi: 10.1186/s40249-015-0090-9] [Medline: 26654247]
46. Liu M, Chan M, Hsu C, Weng C, Yen T, Huang W. Association of uremic pruritus in hemodialysis patients with the number of days of high mean 24-hour particulate matter with a diameter of <2.5 µm. Ther Clin Risk Manag 2017;13:255-262 [FREE Full text] [doi: 10.2147/TCRM.S129133] [Medline: 28260912]
47. Valdes-Rodriguez R, Mollanazar NK, González-Muro J, Nattkemper L, Torres-Alvarez B, López-Esqueda FJ, et al. Itch prevalence and characteristics in a Hispanic geriatric population: A comprehensive study using a standardized itch questionnaire. Acta Derm Venereol 2015 Apr;95(4):417-421 [FREE Full text] [doi: 10.2340/00015555-1968] [Medline: 25203328]
48. Grandhi R, He A, Semenov YR, Kwatra SG. Seasonal variation of itch: A study using real-time data from 2004 to 2016. J Am Acad Dermatol 2017 Mar;76(3):563-564. [doi: 10.1016/j.jaad.2016.09.046] [Medline: 28212763]
49. Bragazzi NL, Watad A, Brigo F, Adawi M, Amital H, Shoenfeld Y. Public health awareness of autoimmune diseases after the death of a celebrity. Clin Rheumatol 2017 Aug;36(8):1911-1917. [doi: 10.1007/s10067-016-3513-5] [Medline: 28000011]
50. Domnich A, Panatto D, Signori A, Lai PL, Gasparini R, Amicizia D. Age-related differences in the accuracy of Web query-based predictions of influenza-like illness. PLoS One 2015;10(5):e0127754 [FREE Full text] [doi: 10.1371/journal.pone.0127754] [Medline: 26011418]


## Abbreviations

**ANOVA:** one-way analysis of variance
**GP:** general practitioner
**OR:** odds ratio
**PM10:** particulate matter 10 micrometers or less in diameter


*Edited by G Eysenbach; submitted 18.02.19; peer-reviewed by E Papadavid, S Kardes; comments to author 08.05.19; revised version received 24.05.19; accepted 04.06.19; published 12.07.19.*

*Please cite as:*
*Tizek L, Schielein M, Rüth M, Ständer S, Pereira MP, Eberlein B, Biedermann T, Zink A*
*Influence of Climate on Google Internet Searches for Pruritus Across 16 German Cities: Retrospective Analysis*
*J Med Internet Res 2019;21(7):e13739*
*URL: http://www.jmir.org/2019/7/e13739/*
*doi:10.2196/13739*
*PMID:31301128*

Original Paper

# Characterizing Swisher Little Cigar–Related Posts on Twitter in 2018: Text Analysis

Jon-Patrick Allem[1], MA, PhD; Sree Priyanka Uppu[2], MS; Tess Boley Cruz[1], PhD; Jennifer B Unger[1], PhD

[1]Keck School of Medicine of USC, Los Angeles, CA, United States

[2]Department of Computer Science, University of Southern California, Los Angeles, CA, United States

**Corresponding Author:**
Jon-Patrick Allem, MA, PhD
Keck School of Medicine of USC
2001 N Soto Street
Los Angeles, CA,
United States
Phone: 1 8586030812
Email: allem@usc.edu

## Abstract

**Background:** Little cigars are growing in popularity in the United States, and Swisher is the market leader. The contexts and experiences associated with the use of Swisher-related products is understudied, but such information is available via publicly available posts on Twitter.

**Objective:** This study aimed to analyze Twitter posts to characterize Twitter users' recent experiences with Swisher-related products.

**Methods:** Twitter posts containing the term "swisher" were analyzed from January 1, 2018, to December 31, 2018. Text classifiers were used to identify topics in posts (n=81,333).

**Results:** The most prevalent topic was *Person Tagging* (mentioning a Twitter account in a post; 32.77%), followed by *Flavors* (eg, Grape and Strawberry; 20.96%) and *Swisher use* (eg, smoke swisher; 17.44%). Additional topics included *Cannabis use* (eg, blunt, roll, and gut swisher; 6.26%), *Appeal* (eg, like Swisher; 5.92%), *Dislike* (eg, posts that showed dissatisfaction with Swisher products; 3.53%), *Purchases* (eg, buy swisher; 1.90%), and *Cigar comparison* (eg, mentions of other cigar products including White-owl and Backwoods; 1.64%).

**Conclusions:** This paper describes common contexts and experiences associated with the use of Swisher little cigars from the population posting on Twitter in 2018. These online messages may have offline consequences for tobacco-related behaviors, indicating the need for countering from public health officials. Findings should inform us about targets for surveillance, policy, and interventions addressing Swisher little cigars as well as communication planning and tobacco product counter messaging on Twitter.

***(J Med Internet Res 2019;21(7):e14398)*** doi:10.2196/14398

**KEYWORDS**

little cigar; cigarillo; Swisher; social media; Twitter; tobacco

## Introduction

Little cigars are growing in popularity in the United States, and Swisher is the market leader [1]. Little cigars deliver substantial nicotine doses and relatively more carbon monoxide than cigarettes [2]. Several factors contributing to the growth in little cigar use have been identified, including their availability in flavors, misperceptions about the harms of use, promotional tactics by the tobacco industry, and lower costs [3,4]. For example, adult smokers have suggested that the lower cost of little cigars (compared to cigarettes) was a reason for initiation and continuation of their use [2]. Internal tobacco industry documents have revealed that tobacco companies engaged in a calculated effort to blur the line between little cigars and cigarettes to increase the appeal to cigarette smokers, and the use of flavors facilitated these efforts [5].

The little cigar consumer marketplace, cultural trends, and tobacco product health policies are constantly evolving [6]. The contexts and experiences associated with little cigar use keep changing, making it important to provide timely information on such issues to inform targets for surveillance, policy, and interventions addressing use of little cigars. Public posts on Twitter can be monitored to quickly capture and describe the context of little cigar use in the words of the people who organically discuss this product. In this way, posts to Twitter can serve as a focus group, offering new insights that may be of importance to tobacco control. Among adults in the United States, Twitter is used by 24% of men, 21% of women, 21% of white individuals, 24% of African Americans, and 25% of Hispanics [7]. In this study, we collected data from Twitter to describe Swisher-related conversations in 2018. Our goal was to describe the public's recent experiences with Swisher, including gaining an understanding of the context in which little cigars are used.

## Methods

### Data Collection

Twitter (twitter.com) posts containing the term "swisher" were obtained from Twitter's streaming application program interface (filtered stream using the Twitter4J library for collecting tweets without gaps in the collection time) from January 1, 2018, to December 31, 2018. The term "swisher" is in line with prior research on little cigars utilizing data from social media [8]. We recorded a total of 111,263 posts during this period.

### Data Processing

To prepare the data for analyses, we excluded non-English tweets, retweets, instances where Swisher was identified as a surname, and tweets from accounts identified as social bots [9], which resulted in a final analytic sample of 81,333 tweets from 57,838 unique users.

In line with prior research [10,11], all tweets in the analytic sample were subjected to basic normalization (eg, lower case all tweets, removal of extra spaces, punctuation between words, and special characters such as brackets), stop word removal (eg, words such as "a" and "the"), lemmatization (eg, breaking down words into their basic form by removing inflections and variants), normalizing mentions of Twitter accounts (eg, @account_name is replaced by @person, which is a common token for all accounts), removal of nonprintable characters (eg, emoticons or symbols in other languages), and removal of hashtags and URLs. All analyses relied on public anonymized data; adhered to the terms and conditions, terms of use, and privacy policies of Twitter; and obtained institutional review board approval from the authors' university. To protect privacy, no tweets were reported verbatim in this report.

### Topic Identification

In line with prior research [10,11], tweets were examined using word counts (frequencies), which included a single word and double-word combinations. From this initial assessment, eight dominant topics were identified by the authors, including *Flavors* (eg, Grape and Strawberry), *Swisher use* (eg, smoke swisher), *Purchases* (eg, buy swisher), *Appeal* (eg, like Swisher), *Cannabis use* (eg, blunt, roll, and gut swisher), *Cigar comparison* (eg, mentions of other cigar products including White-owl and Backwoods), *Dislike* (eg, posts that showed dissatisfaction with Swisher products), and *Person Tagging* (mentioning a Twitter account in a post - @person).

Ultimately, each tweet was classified by checking for the presence of any one of the single words or double-word combinations (n-grams). If a tweet consisted of any of the words associated with a topic, the tweet was classified as part of that topic. In sum, we used a rule-based classification script written in Python where each tweet was checked for the existence of a specified set of n-grams representing a topic [10,11]. For each analysis, we present findings in a confusion matrix where the diagonal line indicates the prevalence of a topic and the off-diagonal lines indicate topic overlap. For example, a hypothetical post such as, "Hey! @person Try Swisher's new grape flavor" could be classified under *Person Tagging* and *Flavors*. The number of posts containing two or more topics would be found at the intersection of the matrix for these topics.

## Results

The eight topics constituted 62.95% of all tweets in the corpus of tweets. The remaining 37.05% of tweets were too varied to be classified into a single topic with meaningful coverage (eg, coverage of each subsequent topic was <1% of total tweets). The most prevalent topic was *PersonTagging* (32.77%; see Multimedia Appendix 1 for common phrases found in this topic), followed by *Flavors* (20.96%) and *Swisher use* (17.45%). *Cannabis use* was the next prevalent topic (6.26%), followed by *Appeal* (5.92%), *Dislike* (3.53%), *Purchases* (1.90%), and *Cigar comparison* (1.64%). *Swisher use* and *Flavors* had the most overlap (12.87%), followed by *Flavors* and *Person Tagging* at (5.57%; Textbox 1 and Table 1).

**Textbox 1.** Topics and common words found in posts along with “Swisher.” These words are meant to provide further context for each topic, are not exhaustive, and are listed in alphabetical order.

**Person tagging:**

@person

**Flavors:**

- Cherry
- Flavor
- Grape
- Mango
- Peach
- Pineapple
- Pumpkin
- Strawberry

**Swisher use:**

- Hit
- Pass
- Puff
- Smoke
- Try

**Appeal:**

- Crave
- Enjoy
- Like
- Love
- Need
- Want

**Dislike:**

- Damn
- Don’t
- Fuck
- No
- Shit

**Cannabis use:**

- Blunt
- Gut
- Marijuana
- Roll
- Weed

**Purchases:**

- Buy
- Bought
- Grab

- Pay

**Cigar comparison:**

- Backwoods
- White-owl

**Table 1.** Prevalence of topics. The diagonal line indicates the prevalence of the eight topics identified. The off-diagonal lines indicate topic overlap. All values are given as n (%).

| | Person tagging | Flavors | Swisher use | Cannabis use | Appeal | Dislike | Purchases | Cigar comparison |
|---|---|---|---|---|---|---|---|---|
| Person tagging | 26,656 (32.77) | —[a] | — | — | — | — | — | — |
| Flavors | 4533 (5.57) | 17,049 (20.96) | — | — | — | — | — | — |
| Swisher use | 3185 (3.92) | 10,464 (12.87) | 14,182 (17.44) | — | — | — | — | — |
| Cannabis use | 1107 (1.36) | 767 (0.94) | 1344 (1.65) | 5088 (6.26) | — | — | — | — |
| Appeal | 815 (1.00) | 855 (1.05) | 386 (0.47) | 171 (0.21) | 4817 (5.92) | — | — | — |
| Dislike | 661 (0.81) | 295 (0.36) | 287 (0.35) | 72 (0.09) | 25 (0.03) | 2869 (3.53) | — | — |
| Purchases | 310 (0.38) | 393 (0.48) | 333 (0.41) | 134 (0.16) | 20 (0.02) | 40 (0.05) | 1542 (1.90) | — |
| Cigar comparison | 301 (0.37) | 296 (0.36) | 344 (0.42) | 82 (0.10) | 34 (0.04) | 159 (0.20) | 20 (0.02) | 1332 (1.64) |

[a]Not applicable.

## Discussion

### Principal Findings

The topics identified in this study of Swisher-related posts on Twitter in 2018 provide several insights about the public's recent experience with little cigars. *Flavors* was a common topic in this study, similar to earlier Twitter-based studies focused on tobacco products like JUUL [10] and hookah [11]. Content analysis of little cigar–related videos posted to YouTube demonstrated that common themes included their candy flavors [12]. Flavors were cited as important reasons for use of little cigars and cigarillos among a national probability sample of US adults in 2014 [3]. Research also suggests that the perception of risk of flavors in little cigars is related to use of these products, specifically, the perception of less risk [13]. Additionally, more than two-fifths of US middle and high school smokers report using flavored little cigars or flavored cigarettes [14]. Taken together, flavors in little cigars may be considered a priority area for federal regulation to reduce appeal of these little cigars to consumers across age groups and to provide uniform restrictions that make it difficult for distributors and consumers to work around local flavor restrictions.

In a previous assessment of public tweets related to little cigars, Stepp and colleagues found that posts often expressed affiliation for specific brands (Swisher Sweets and Black & Mild) as well as reporting smoking activity [15]. In this study, *Purchases, Swisher use, Appeal, Cigar comparison,* and *Person tagging* were common topics in the data. These results together suggest that Twitter users are talking about their smoking activities; comparing and contrasting their preferences for brands; and directly communicating with their followers about such purchases, preferences, and activities. These online messages may have offline consequences on tobacco-related behaviors [16], suggesting that such messages need countering from public health officials.

*Cannabis use* was a common topic in this study. Prior analysis of Instagram posts revealed that Swisher little cigars were often gutted and filled with cannabis [8]. Little cigars may be at the intersection of nicotine and cannabis use, raising major public health concerns, including increased risk of transition between cannabis and tobacco [17], high frequency of use [18], and addiction to tobacco [18]. For example, ever use of marijuana has been shown to be a predictor of initiation of regular little cigar and cigarillo use among US young adults (ages of 18-34 years) [19].

Prior studies rarely reported that Twitter users voiced dissatisfaction with tobacco [10,11]. However, *Dislike* was a topic identified in this study. Although this topic did not strongly overlap with any of the other topics to indicate further context, it suggests that there are general forms of complaints that could be amplified by public health practitioners to discourage tobacco use, in general, or the uptake of little cigar use, in particular.

### Limitations

This study focused on posts on Twitter, and its findings may not generalize to other social media platforms. Data collection relied on Twitter's streaming application program interface, which prevented collection of tweets from private accounts. Therefore, our findings may not represent the attitudes and behaviors from individuals with private accounts. The posts analyzed in this study were collected from a 12-month period and may not be generalizable to other time periods. Although only one little cigar brand was the focus of this study, Swisher is the little cigar market leader and has been the focus of prior research [8].

## Conclusions

This paper described the common contexts and experiences associated with Twitter discussions about Swisher little cigars in 2018. The predominant conversation topic contained some form of interpersonal communication (person tagging) that would capture the social nature of the posts, flag another person, or refer to them as a notable source. Flavors, the second most common category, are the main characteristic that could be regulated in the future to reduce appeal. Findings should inform targets for surveillance, policy, and interventions addressing little cigars as well as communication planning and tobacco product counter messaging on Twitter.

## Acknowledgments

Research reported in this publication was supported by Grant #U54 CA180905 from the National Cancer Institute and the Food and Drug Administration (FDA) Center for Tobacco Products. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH) or FDA.

## Authors' Contributions

JA and SU conceived of the study and analyzed the data. JA drafted the initial manuscript. SU, TC, and JU revised the manuscript for important intellectual content and approved the final manuscript. JU and TC received funding for the study.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Word cloud of common phrases found in Person Tagging.

[PDF File (Adobe PDF File), 7MB - jmir_v21i7e14398_app1.pdf ]

## References

1. Centers for Disease Control and Prevention. 2018 Aug 01. Tobacco brand preferences URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/tobacco_industry/brand_preference [accessed 2019-03-26]
2. Pickworth WB, Rosenberry ZR, Koszowski B. Toxicant exposure from smoking a little cigar: further support for product regulation. Tob Control 2017 Dec;26(3):269-276. [doi: 10.1136/tobaccocontrol-2015-052633] [Medline: 27122063]
3. Nyman AL, Sterling KL, Weaver SR, Majeed BA, Eriksen MP. Little Cigars and Cigarillos: Users, Perceptions, and Reasons for Use. Tob Regul Sci 2016 Jul;2(3):239-251 [FREE Full text] [doi: 10.18001/TRS.2.3.4] [Medline: 27413772]
4. Sterling K, Majeed B, Nyman A, Eriksen M. Risk Perceptions of Little Cigar and Cigarillo Smoking Among Adult Current Cigarette Smokers. Nicotine Tob Res 2017 Nov 01;19(11):1351-1358 [FREE Full text] [doi: 10.1093/ntr/ntw244] [Medline: 27659275]
5. Kostygina G, Glantz S, Ling P. Tobacco industry use of flavours to recruit new users of little cigars and cigarillos. Tob Control 2016;25(1):66-74. [Medline: 25354674]
6. Delnevo CD, Hrywna M, Giovenco DP, Miller Lo EJ, O'Connor RJ. Close, but no cigar: certain cigars are pseudo-cigarettes designed to evade regulation. Tob Control 2017 Dec;26(3):349-354 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-052935] [Medline: 27220622]
7. Perrin A, Anderson M. Pew Research Center. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018 URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-05-08]
8. Allem JP, Escobedo P, Chu KH, Boley Cruz T, Unger JB. Images of Little Cigars and Cigarillos on Instagram Identified by the Hashtag #swisher: Thematic Analysis. J Med Internet Res 2017 Dec 14;19(7):e255 [FREE Full text] [doi: 10.2196/jmir.7634] [Medline: 28710057]
9. Allem JP, Ferrara E. The Importance of Debiasing Social Media Data to Better Understand E-Cigarette-Related Attitudes and Behaviors. J Med Internet Res 2016 Dec 09;18(8):e219 [FREE Full text] [doi: 10.2196/jmir.6185] [Medline: 27507563]
10. Allem JP, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. Drug Alcohol Depend 2018 Dec 01;190:1-5. [doi: 10.1016/j.drugalcdep.2018.05.018] [Medline: 29958115]
11. Allem JP, Dharmapuri L, Leventhal AM, Unger JB, Cruz TB. Hookah-Related Posts to Twitter From 2017 to 2018: Thematic Analysis. J Med Internet Res 2018 Dec 19;20(11):e11669 [FREE Full text] [doi: 10.2196/11669] [Medline: 30455162]
12. Richardson A, Vallone DM. YouTube: a promotional vehicle for little cigars and cigarillos? Tob Control 2014 Jan;23(1):21-26. [doi: 10.1136/tobaccocontrol-2012-050562] [Medline: 23047887]
13. Nyman AL, Sterling KL, Majeed BA, Jones DM, Eriksen MP. Flavors and Risk: Perceptions of Flavors in Little Cigars and Cigarillos Among U.S. Adults, 2015. Nicotine Tob Res 2018 Aug 14;20(9):1055-1061. [doi: 10.1093/ntr/ntx153] [Medline: 28666353]

14. King BA, Tynan MA, Dube SR, Arrazola R. Flavored-little-cigar and flavored-cigarette use among U.S. middle and high school students. J Adolesc Health 2014 Jan;54(1):40-46 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.07.033] [Medline: 24161587]
15. Step MM, Bracken CC, Trapl ES, Flocke SA. User and Content Characteristics of Public Tweets Referencing Little Cigars. Am J Health Behav 2016 Jan;40(1):38-47. [doi: 10.5993/AJHB.40.1.5] [Medline: 26685812]
16. Unger JB, Urman R, Cruz TB, Majmundar A, Barrington-Trimis J, Pentz MA, et al. Talking about tobacco on Twitter is associated with tobacco product use. Prev Med 2018 Dec;114:54-56. [doi: 10.1016/j.ypmed.2018.06.006] [Medline: 29898418]
17. Audrain-McGovern J, Stone MD, Barrington-Trimis J, Unger JB, Leventhal AM. Adolescent E-Cigarette, Hookah, and Conventional Cigarette Use and Subsequent Marijuana Use. Pediatrics 2018 Dec;142(3):e20173616. [doi: 10.1542/peds.2017-3616] [Medline: 30082450]
18. Volkow ND, Baler RD, Compton WM, Weiss SR. Adverse health effects of marijuana use. N Engl J Med 2014 Jun 05;370(23):2219-2227 [FREE Full text] [doi: 10.1056/NEJMra1402309] [Medline: 24897085]
19. Cohn AM, Johnson AL, Fryer CS, Villanti AC. Marijuana use predicts onset of current little cigar use in a national sample of US young adults. Drug Alcohol Depend 2018 Dec 01;190:235-241. [doi: 10.1016/j.drugalcdep.2018.01.020] [Medline: 30064060]

*Edited by G Eysenbach; submitted 15.04.19; peer-reviewed by K Rehberg, L Laestadius; comments to author 24.05.19; revised version received 03.06.19; accepted 18.06.19; published 19.07.19.*

*Please cite as:*
*Allem JP, Uppu SP, Boley Cruz T, Unger JB*
*Characterizing Swisher Little Cigar–Related Posts on Twitter in 2018: Text Analysis*
*J Med Internet Res 2019;21(7):e14398*
*URL: http://www.jmir.org/2019/7/e14398/*
*doi:10.2196/14398*
*PMID:31325291*

Original Paper

# Perceptions of Information and Communication Technology as Support for Family Members of Persons With Heart Failure: Qualitative Study

Hanna Allemann[1*], RN, MSc; Ingela Thylén[1,2*], RN, PhD; Susanna Ågren[1,3*], RN, PhD; Maria Liljeroos[1,4*], RN, PhD; Anna Strömberg[1,2*], RN, PhD

[1]Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

[2]Department of Cardiology, Linköping University, Linköping, Sweden

[3]Department of Cardiothoracic Surgery, Linköping University, Linköping, Sweden

[4]Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden

[*]all authors contributed equally

**Corresponding Author:**
Hanna Allemann, RN, MSc
Department of Medical and Health Sciences
Linköping University
Division of Nursing, Campus US
Linköping, 58183
Sweden
Phone: 46 13 28 67 76
Email: hanna.allemann@liu.se

## Abstract

**Background:** Heart failure (HF) affects not only the person diagnosed with the syndrome but also family members, who often have the role of informal carers. The needs of these carers are not always met, and information and communications technology (ICT) could have the potential to support them in their everyday life. However, knowledge is lacking about how family members perceive ICT and see opportunities for this technology to support them.

**Objective:** The aim of this study was to explore the perceptions of ICT solutions as supportive aids among family members of persons with HF.

**Methods:** A qualitative design was applied. A total of 8 focus groups, comprising 23 family members of persons affected by HF, were conducted between March 2015 and January 2017. Participants were recruited from 1 hospital in Sweden. A purposeful sampling strategy was used to find family members of persons with symptomatic HF from diverse backgrounds. Data were analyzed using qualitative content analysis.

**Results:** The analysis revealed 4 categories and 9 subcategories. The first category, about how ICT could provide relevant support, included descriptions of how ICT could be used for communication with health care personnel, for information and communication retrieval, plus opportunities to interact with persons in similar life situations and to share support with peers and extended family. The second category, about how ICT could provide access, entailed how ICT could offer solutions not bound by time or place and how it could be both timely and adaptable to different life situations. ICT could also provide an arena for family members to which they might not otherwise have had access. The third category concerned how ICT could be too impersonal and how it could entail limited personal interaction and individualization, which could lead to concerns about usability. It was emphasized that ICT could not replace physical meetings. The fourth category considered how ICT could be out of scope, reflecting the fact that some family members were generally uninterested in ICT and had difficulties envisioning how it could be used for support. It was also discussed as more of a solution for the future.

**Conclusions:** Family members described multiple uses for ICT and agreed that ICT could provide access to relevant sources of information from which family members could potentially exchange support. ICT was also considered to have its limitations and was out of scope for some but with expected use in the future. Even though some family members seemed hesitant about ICT solutions in general, this might not mean they are unreceptive to suggestions about their usage in, for example, health care. Thus, a variety of factors should be considered to facilitate future implementations of ICT tools in clinical practice.

*(J Med Internet Res 2019;21(7):e13521)* doi:10.2196/13521

**KEYWORDS**

family; caregivers; telemedicine; perception; heart failure; social support; focus groups; qualitative research

## Introduction

### Heart Failure

Heart failure (HF) is a syndrome affecting approximately 1% to 2% of the adult population in most countries of the world. The prevalence increases with age, rising to over 10% in those older than 75 years [1]. Worldwide, HF not only causes considerable suffering for the persons with the condition but also takes up considerable societal economic resources [2]. Having HF affects health through physical symptoms such as breathlessness, fatigue, and swollen extremities [1] and is also associated with psychosocial distress [3,4], for example, depression and anxiety.

### Informal Carers to Persons With Heart Failure

HF deteriorates over time and is unpredictable in its course [1]. HF also impinges on family, friends, and significant others [5,6], especially when the ill person experiences more severe symptoms [7]. Family members often shoulder the role of informal carers, supporting persons with HF both practically and emotionally [5,6,8] and are instrumental in supporting self-care [9] as well as in helping the ill person navigate the health care system [6,9,10].

Informal caregiving can be both rewarding [11,12] and straining [12-14]. Positive experiences of caregiving might include the following: an increase in self-esteem, a feeling of pride [11], and an intensification of the relationship with the loved one suffering from HF [11,14]. Possible negative effects of caregiving are multifaceted [5], ranging from physical, emotional, and social burdens [11] to isolation [5,15]. Informal carers have described feelings of always having to be *on stand-by* [6,10,14] as well as struggling to balance taking care of the ill person, themselves, and maintaining their family and working life [6,10]. It has been established that the needs of carers are not always accommodated [6,16]. In terms of what caregivers do, they describe a need for information and support from health care that is lacking [5,6]. Being a family member or an informal carer of a person with HF could be challenging because of the illness trajectory, where shifts in day-to-day health [6], together with symptoms that will undoubtedly worsen over time, create an unpredictable life situation. This may apply both for those suffering from the syndrome and for family, friends, and significant others. Family members who care for persons with HF are also vulnerable in that they are often older and have health issues of their own [16].

### Information and Communications Technology to Support Informal Carers

Information and communications technology (ICT) could facilitate caregiving, support carers in managing their life situation [17], reduce carers' burden [18,19], and possibly strengthen self-efficacy [19]. According to a recent scoping review aimed at evaluating Web interventions for informal carers of older persons, carers found ICT feasible and usable. One conclusion was that it is important to adapt interventions to caregivers' different and changeable needs. However, none of the interventions focused on HF [19]. Providing relevant support for those close to the ill person has the potential to bring positive effects for the family member acting as informal carer as well as the person with HF; however, the use of ICT is relatively unexplored within HF [6].

### Objectives

Knowledge already exists about what needs caregivers might have in relation to the carer role/situation, but there is an apprehended knowledge gap with regard to how family members of persons with HF value and perceive ICT in their everyday life in relation to these needs. The insight into how ICT is perceived by family members could contribute important knowledge that could be used in developing interventions to support informal carers as well as supporting the implementation of ICT clinically. Therefore, the aim of this study was to explore the perceptions of ICT solutions as supportive aids among family members of persons with HF.

## Methods

### Study Design

This study had a qualitative design. Focus group discussions with family members of persons afflicted with HF were performed, and data were analyzed using qualitative content analysis as described by Elo and Kyngäs [20]. The study complies with the Declaration of Helsinki and was approved by the regional ethical review board in Linköping (ref #2015/55-32).

### Participants and Sampling

A purposeful sampling strategy [21] was used with the intention of finding family members from diverse backgrounds, for example, diversity in gender, cultural background, age, and relationship to the person with HF. Initially, patients from an outpatient nurse-led HF clinic at 1 university hospital in Sweden were approached when visiting the clinic, or otherwise contacted by a research nurse. If the patient agreed, he or she was asked to select a family member involved in their care and self-care. Only family members who were able to participate in a focus group interview could be invited, for example, family members who spoke Swedish and had no severe cognitive or hearing impairment. An information letter was sent out or handed to the family member, followed by telephone contact. If the family member agreed to join, a written consent form was signed.

### Data Collection

Before initiating the focus group discussion, participants answered a self-reported questionnaire, developed for this study, about their demographic and clinical characteristics and use of the internet. Participants were informed about the aim and goals of the focus groups. The setup of the focus group discussions was guided by Krueger and Casey [22], and the interview guide had opening, introductory, transition, key, and ending questions to build a logical flow, while also focusing the discussions on support and technology (see Multimedia Appendix 1). Opening and introductory questions were about creating a safe environment. In transition and key questions, participants were asked to reflect on support and to discuss their own experiences of supporting practically, emotionally, informationally, and through confirmation. They were also asked to describe how they cared for themselves in their everyday life. The final key question was a direct question concerning ICT: *What do you imagine internet technology could support you with in your daily life as a family member of someone with HF?* Probing questions, for example, *could you provide concrete examples, could you expand on that, what could help you to provide this care*, were asked to deepen discussions. A pilot focus group was performed to test the procedure, and as no significant changes were made, the pilot group was later included in the analysis.

A total of 2 researchers attended each group and tape-recorded every session, which took place in a secluded meeting room at the university. All the researchers involved in the focus groups (HA, ML, SÅ, and AS) are registered nurses, and 3 out of 4 (ML, SÅ, and AS) have joint positions combining research and clinical work. A moderator (HA or ML) introduced the subjects and questions to be discussed and ensured that all the participants were given a chance to speak. The observer (SÅ or AS) took notes about the discussion and interactions and was able to ask additional questions. The observer was also responsible for producing a summary at the end that the participants could accept, correct, or expand.

A Microsoft PowerPoint presentation with the open-ended interview questions and examples of, for example, existing internet technology solutions were projected during the discussions. These examples included Web meetings, apps, seeking information via the web, and chatting with others. The presentation also included a definition of what a digital service could be about (accessing information without being bound to a certain place and including pictures, films, and text). Participants were reminded to reflect on their role as family members of someone with HF, as opposed to just reflecting more generally on questions. The focus group discussions were followed by a debrief between the moderator and observer to compare understanding of, for example, interactions among participants. All the focus groups were performed between March 2015 and January 2017.

## Analysis

Interviews were transcribed verbatim. To find relevant content and become immersed in the material, the tape-recordings were listened to, and transcriptions read, several times. Afterward, the *unit of analysis* was identified by 2 of the authors (HA and ML). The unit of analysis meant text concerning ICT. The first author then read the unit of analysis several times and started the open coding, making notes and headings in the margins, with the codes then sorted in a coding sheet. The codes and notes were then used to freely generate categories, which in turn were sorted and merged into fewer, more overarching categories. The first author was responsible for developing codes and categories, and another (ML) author verified and suggested changes. This was an iterative and nonlinear process, moving back and forth between part and whole, until no new codes or categories merged. During the process, the authors were careful to stay close to what the participants actually said, while also keeping in mind that an abstraction occurred through the different phases of the analysis. When categories were defined, these and the coding scheme were reviewed by AS and IT for agreement or discussion until a consensus was reached (Table 1).

**Table 1.** Analytical process of focus group data.

| Excerpt from unit of analysis | Open coding | Subcategory | Category |
|---|---|---|---|
| "...I think it would be great, it [interacting with other relatives via ICT[a]] would often make it easier. [But] then you realize that everyone is different and a response given to someone may not apply to my husband, but I think that a lot of this with heart failure is...about calming down a bit..." | Even if everyone is different, it might calm worry, to hear other people's stories. | For exchange with peers and external family | ICT—Providing possibilities for relevant support |
| "[Discussing having contact with health care personnel via the internet]...those who respond to this could be anywhere, as long as they have the knowledge [Other participant: yes, that's right], so you can get a quick response to see what the possible next step may be [Other participant: mm]..." | ICT is not tied to a specific location; ICT used to write questions to health care professionals and get quick answers. | Unbound by time and place, with endless possibilities | ICT—Providing access |
| "...so I´d rather meet face-to-face, I want to see that person. I want to see their responses when I ask questions. I think one interprets facial expressions [several other participants confirm] too..." | Wants to see the person when talking to them to be able to read body language. | Physical meetings are irreplaceable | ICT—Being too impersonal |
| "...it's too new...we haven't experienced enough really..." | It is too new [the HF[b] diagnosis] to be able to see the need for ICT. | Difficult to visualize | ICT—Being out of scope |

[a]ICT: information and communications technology.

[b]HF: heart failure.

## Results

### Characteristics of Participants

Participants had diverse backgrounds in terms of, for example, education and occupational status, whereas being more homogenous in terms of, for example, living status and relationship to the person with HF (Table 2). Almost all participants reported having access to the internet, and many used the Web daily. Only 4 had sought information about being a family member or an informal carer (Table 3). A total of 8 focus groups involving 23 family members of persons with HF were conducted. The recordings lasted 55 to 121 min (mean, 91 min). Analysis resulted in 4 categories and 9 subcategories (Table 4).

**Table 2.** Self-reported characteristics of participants (N=23).

| Characteristics | Value |
|---|---|
| Age (years), median (range)[a] | 63 (26-85) |
| **Gender, n** | |
| Women | 18 |
| **Education, n** | |
| Compulsory school | 3 |
| Upper secondary school | 8 |
| University | 8 |
| Other | 4 |
| **Main occupation, n** | |
| Employed | 8 |
| Self-employed | 1 |
| Student | 1 |
| Retired | 12 |
| On sick-leave | 1 |
| Living with a partner, n | 23 |
| Living with children, n | 6 |
| **Health problems, n[a]** | |
| Diabetes, yes | 1 |
| High blood pressure, yes | 7 |
| Rheumatic disease, yes | 1 |
| Stroke, yes | 1 |
| Lung disease, yes | 1 |
| Atrial fibrillation, yes | 1 |
| Myocardial infarction, yes | 1 |
| Other disease, yes | 2 |
| **Relationship to person with heart failure, n** | |
| Married/partner | 22 |
| Child | 1 |

[a]Some missing values.

**Table 3.** Self-reported internet use among participants (N=23).

| Internet use | Frequency, n |
| --- | --- |
| Access to internet at home, yes[a] | 21 |
| Access internet away from home, yes[a] | 15 |
| **How often has the internet been used in the last 3 months[a]** | |
| Almost every day | 17 |
| At least once a week | 3 |
| Less than once a week | 1 |
| Do not use | 1 |
| **What was the internet used for in the last 3 months** | |
| Access to internet banking[a] | 20 |
| Email[a] | 19 |
| Seeking information[a] | 18 |
| Access to news sites[a] | 17 |
| Travel services[a] | 16 |
| Seeking illness-related information[a] | 15 |
| Using apps, music, films, playing games | 14 |
| Selling/buying goods or services[a] | 13 |
| Social media[a] | 12 |
| Video call[a] | 10 |
| Uploading self-made material | 8 |
| Listening to internet radio | 7 |
| Booking appointments for health services | 5 |
| Seeking information about informal caregiving[a] | 4 |
| Creating a website or blog | 4 |
| Playing online games with others | 3 |

[a]One missing value.

**Table 4.** Categories and subcategories from analysis of focus group discussions with family members of persons with heart failure, along with the number of focus groups in which the subcategory occurred.

| Category and subcategory | Focus groups, n |
|---|---|
| **Providing possibilities for relevant support** | |
| Interaction with health care personnel | 7 |
| Information and confirmation retrieval | 6 |
| Exchange with peers and external family | 7 |
| **Providing access** | |
| Unbound by time and place, with endless possibilities | 7 |
| Arena for family members | 5 |
| **Being too impersonal** | |
| Apprehensions about usability | 7 |
| Physical meetings are irreplaceable | 5 |
| **Being out of scope** | |
| Difficult to visualize | 5 |
| Something for the future | 5 |

## Information and Communications Technology: Providing Possibilities for Relevant Support

Groups discussed several uses of ICT, including the possibility of communicating with professionals, retrieving information—either to learn about HF/caregiving or to get confirmation of their own thoughts or actions relevant to them. The opportunity to interact with persons in similar life situations, to find and share support, was another subject raised in discussions.

### *Interaction With Health Care Personnel*

Almost all the groups discussed the possibility of ICT providing a forum for interacting with health care personnel and receiving individual support. Different aspects of interaction via ICT came up, and being able to ask questions and to receive specific advice was mentioned. Other aspects conveyed were the need for those giving advice or answering questions to have knowledge about HF specifically:

> *Participant: [discussing communicating with health care] ...If there was a nurse or a physician or a psychologist...who you could ask regular questions and get answers from. I would really appreciate that...*
>
> *Moderator: ...someone who has specific knowledge...*
>
> *Participant: Yes...*

Being able to put questions to health care personnel specializing in HF was discussed as a way of avoiding information that you do not want or feel the need for. Another aspect raised was the importance of straightforward information that did not hold back on details for fear of causing family members or patients worry.

Web meetings led by professional staff and social media forums were possible arenas for the interaction. It was pointed out that it was important to maintain the privacy of users´ details when arranging closed group sessions on the internet. The importance of Web security was raised in some groups. This entailed both the importance of finding valid sources of information and using secure solutions for communication with health care personnel or other family members through Web-based solutions.

A couple of groups discussed the use of ICT for conveying contacts that could be important to the family member, and it was pointed out that these contacts should be relevant in relation to having HF in the family. One participant put it like this:

> *We sometimes experience frustration about not knowing where to turn with perhaps quite a simple question.*

Another aspect of the interaction with health care personnel was the possibility of using ICT for follow-up of the person with HF. This could potentially unburden the family member. Other positive examples of helpful solutions were remote rhythm monitoring of the person with HF living with an implanted cardiac defibrillator, and the possibility of following up on hospital visits to deal with questions arising after a visit to the physician, for example.

### *Information and Confirmation Retrieval*

According to the family members, ICT offered a way to retrieve information about HF-specific matters, and several groups discussed already having used this possibility. Several practical examples of HF-specific information for family members were given. The content presented through ICT could entail information about symptoms of HF, the consequences of symptoms, and warning signs—what to look out for in the ill person: (information about)...*warning signs* (relating to HF) *that you should look out for, that could be of help*. Information concerning care trajectories was also mentioned as well as information on heredity. Another topic concerned cardiac-related anxiety and how to separate this from actual symptoms of HF, and how to tackle this as a family member. ICT could also be used to convey concrete information and support concerning what family members might encounter when the person with HF is in hospital. A wish to read about HF in younger persons—about how it is possible to have a good life even though illness is present in the family—was also conveyed. Yet another aspect was that it would be supportive, if advice about what to eat, were translated into concrete menus. Family members also wished to receive discharge notes aimed specifically at them. ICT could also involve suggestions about questions to ask during visits to health care institutions.

The possibility of ICT being used to organize information in a comprehensible and easy-access format was also mentioned. It was suggested that the need for information was related to the symptoms of the ill person. When the family member was first diagnosed, it was important to read about medicines and side effects, but when the person with HF was in a stable phase, this need did not seem equally apparent.

ICT could also be used to receive confirmation and support decision making—information through ICT could help determine whether to contact health care or not:

> *...when something happens at home then you go to the internet and look [other participant hums]...before deciding to call 1177 [a national telenursing service number] [laughs] you could do a bit of research on your own...*

Confirmation of the family member´s own actions and support in relation to the ill person could help to reassure the family member.

Some spontaneously discussed how information and confirmation could be delivered through text on a Web page, through a recorded lecture, a frequently asked questions section, and through social media forums. It was also stressed that information on the Web should be presented so that everyone could understand, meaning that not too many medical terms should be used.

### *Exchange With Peers and External Family*

ICT could be useful for contact with other family members of persons with HF. This seemed to have social, practical, and emotional functions, and interacting with other family members could entail a learning experience as expressed by this participant:

> *...you can read the posts [on the internet] and maybe learn something from it. Because it’s a bit difficult and you will never have learnt enough as a next of kin...*

Contact with other family members could also calm worry and support problem solving. Receiving advice from others, for example, on what to consider in specific situations, as well as the possibility of sharing one’s own experiences, was discussed. Connecting with other family members through ICT seemed to be a way of recognizing oneself in others.

The opportunity to connect with others in similar life situations could also be valuable for the persons with HF, and it could be important to have separate groups for family members and persons with HF as well as mixed groups:

> *[Answering a question from the moderator about who a social media group would be aimed at, the patient or the family member?]...I think, both, but at the same time you also need to be able to talk about them [referring to their family member with HF], even though they don’t always know...so both...because they also need to talk about it [referring to person with HF].*

Also mentioned was the fact that it could be easier to *talk via the Web* if you had met the person in real life first, or that connecting with others through ICT could be a prerequisite for seeing each other in real life later.

Chat groups and social media forums were raised as possible Web-based solutions for interaction with other family members.

A few groups also discussed how ICT could be used for reaching or keeping in contact with extended family. The use of social media could, for example, mean that one did not have to have personal contact with everyone in the family in times of crisis. By posting information on social media, you could relieve yourself of having to answer questions from extended family, thus enabling you to focus on the ill person. On the other hand, the possibility of maintaining personal contact was also described as a strength of ICT.

## Information and Communications Technology: Providing Access

ICT could offer solutions accessible from anywhere and whenever, and ICT could be both timely and adaptable to different life situations. ICT could also provide an arena for family members that they might not have had access to otherwise.

### *Unbound by Time and Place, With Endless Possibilities*

Groups discussed how ICT allowed the option of adjusting times when information could be accessed and provided via the Web. This was mentioned as a pro for, for example, family members who worked and for those with little spare time during regular working hours.

The groups discussed how ICT was unbound to a specific place when accessing or providing information, something that was also mentioned as being favorable for health care personnel. Remote solutions, for example, checking blood pressure, weight, and symptoms and having an initial assessment done via an internet-based communication tool were mentioned as one possibility. ICT was also stated as a solution that would bring health care closer to the patient and family member:

> *I think that taking the doctor or the care to him [the patient, through ICT] would be very useful and then I could feel that I can relax a little...*

It was also suggested that ICT could provide rapid answers to questions and concerns and be something that could give access to *everything*. When discussing this, it was not always explicitly expressed *how* ICT had a specific use for the family members. These statements seemed to be more about showing an understanding that “you have the world in your pocket” when having access to the internet, and if you know where to look and have the knowledge to use computers, you have access to everything:

> *...if you have an internet connection, you actually have all the possibilities in the world...[another participant confirms].*

### Arena for Family Members

ICT could provide an arena for the family member to take part of information that concerns the person with HF but also for addressing one's own questions. Family members did not always feel they were able to take up space in health care consultations:

> *It concerns him [referring to the ill person]...so you don't have that space as family at all...*

Some discussed opportunities to ask questions without their ill family member present; however, others thought it was important that both the person with HF and the family member had access to the same information.

ICT could increase their participation in caring for the ill person. Receiving information about the person with HF while he or she was hospitalized—even when not being able to be present—could ease involvement:

> *...I would like to be...a little bit more involved. Because I was not physically here [at the hospital] at these times...one could have received documentation [concerning status]...[other participants confirm]. It would have been very easy, either by e-mail or by regular mail.*

## Information and Communications Technology: Being Too Impersonal

ICT was perceived as entailing limited personal interaction and individualization, which could cause mistrust of information and could cause a feeling of ICT being less usable. It was made clear in several discussions that ICT could not replace physical meetings.

### Apprehensions About Usability

It could be hard to recognize oneself in information provided through ICT, as expressed by this participant:

> *...it rarely fits you directly [discussing Web-based information], or maybe you imagine that it does, but it doesn´t really...*

Groups discussed that ICT could not be used for everything, for example, for support in acute situations or for dealing with worry:

> *...well, a disadvantage of the chat function is that I can't convey the worried feeling I might have...*

Concerning acute situations, an app for support in the event of a cardiac arrest was mentioned. Even though this app was recognized as a valid alternative, it was also implied that it was not necessarily trusted.

Some discussed how information on the internet could provide answers to *some* questions but that it was seen as being too based on facts and not personal enough for everything. This indicated a need for personalized and tailored information and was something that could get in the way of ICT being a usable aid for support.

Other aspects were that ICT was not personalized enough. Some family members also discussed how seeking information through ICT could cause worry and lead to imagining illnesses or problems that were not there:

> *...I think that you eventually only get sicker and sicker if you read everything, and more worried too.*

Even though groups discussed the supportive effect of communicating with other family members through ICT, some were hesitant about the idea of connecting with other family members via the Web. Some questioned meeting *strangers* via the Web, and others discussed how communicating through ICT could mean that discussions might *derail* and that some persons could *show too much of themselves.* In connection to this, it was mentioned that contact with other family members via ICT needed to be organized to avoid this.

### Physical Meetings Are Irreplaceable

ICT could not replace physical meetings. The reasons for this were already having the necessary contacts with health care personnel in real life or not feeling a need for ICT. Some just stated that they prefer to see someone "eye to eye" instead of receiving information via ICT:

> *I don't want information via the internet, I want real information.*

Not allowing body language to be read was given as a disadvantage of ICT. Another reason for preferring physical meetings was that ICT as an aid would not be able to relieve worry.

## Information and Communications Technology: Being Out of Scope

ICT was described as being out of scope, and some were generally uninterested or had less faith in ICT. They did not know much about ICT and were less interested in using it. For some, it seemed hard to envision what ICT could be of use for, in relation to being a family member of someone with HF, and it was thought of as more of a solution for the future.

### Difficult to Visualize

Groups expressed a difficulty in envisioning how ICT could support them in their everyday life. This became clear when groups were approached with questions concerning ICT during the focus group discussion but stated that they did not think they could contribute:

> *I can't think of anything right now, because my first thought was that the internet...couldn't help me with anything...*

It could be that being a family member of someone newly diagnosed with HF also made visualization difficult. However, when presented with examples of what ICT could be about during the focus group discussions, it seemed that some family members took inspiration and incorporated these into their discussions.

### Something for the Future

ICT was considered in some groups to be something suited more for the future and for coming generations but also as something that could be useful if the ill person's health deteriorated. Groups discussed how coming generations might be better equipped to use these kinds of solutions, and it was mentioned that younger generations already had ICT as a natural part of their daily life. Some groups discussed the significance of age and how it could be that they were too old for ICT:

> *One can say that I belong to the older generation, and it is more our children and grandchildren who will use it [referring to using a computer] and get more benefit from it I would think...so in the future...*

Not everyone agreed, and in response to one family member expressing being too old for ICT, another participant responded by pointing out that:

> *You're never "too old to learn something new."*

# Discussion

## Principal Findings

The analysis explored perceptions of ICT solutions as supportive aids among family members of persons with HF. Overall, ICT was seen as having a broad range of uses, while also involving—or lacking—content or qualities that might make ICT less useful for family members of persons with HF.

## Comparison With Prior Work

Participants suggested that ICT could be used for interacting with health care personnel, to give and receive support from other family members, and for communication with the extended family. The interaction with health care personnel was considered to be useful for receiving personal advice from knowledgeable experts, whereas interaction with other family members seemed to be about recognizing oneself in others—and receiving emotional support and advice. The importance of being able to communicate with others in similar situation [23] and health care personnel [24] is reflected in other studies, and connecting with other carers via the Web has been noted as being just as helpful as seeing someone physically [19]. Although our study suggests that for some carers, ICT could not replace physical meetings and could even cause worry, communication with both health care professionals and peers via the Web seemed to be of interest.

In developing support through ICT, there could be a challenge in balancing relevant information with specific information that will not evoke negative feelings. Some participants stated that seeking information on the internet could have negative effects, such as causing anxiety and worry. For some, it was considered important to have a personal contact with the health care personnel to obtain individualized information, whereas others seemed to want both personalized and general information. The need for both types of information mirrors other findings [25]. The need for different kinds of information might also relate to different coping styles, with coping strategies such as being a *monitor* or being *a blunter* impacting on what is considered relevant. Being a monitor or a blunter in relation to use of information has been studied previously [26-28]. It could be that a person who is mainly characterized as a *blunter* avoids threatening information [26] and, therefore, might not seek information via the Web. For those labeled a *monitor* (wanting and scanning for information while also being more anxious) [26], information delivered through ICT could cause added stress if not carefully worded. Factors such as these could be of importance when producing Web based informational material. Questions such as—whom will it actually reach and how could the information affect those reading it, needs to be addressed. When producing material, it may also be worth considering adapting information to the relationship that the family member has to the ill person [29]. In this study, we almost exclusively had discussions with spouses and it might be that, for example, siblings or friends may suggest other uses of ICT.

In our study, some family members felt that ICT could not replace physical meetings, for example, because it hindered the reading of body language and also as it could be difficult to convey worry. This might not necessarily reflect a limitation concerning ICT but rather the fact that it was hard for some family members to visualize how ICT could be a support. In a study concerning ICT support for carers, the conclusion was that when the elderly carers received support from the nurses, they were more likely to use the internet-based intervention [15]. It might be that the initial reaction to suggestions about using ICT for support is not necessarily unsusceptible to influence, and it could therefore be important to educate and support carers in how to operate ICT when implementing these kinds of interventions.

Almost all the participants in the focus groups had access to and used the internet on a daily basis. Many also reported having access away from home (eg, via a mobile phone). Even so, not all expressed an interest in the use of technology in their everyday life. Some said they were too old for ICT, and some expressed a lack of interest or even mistrust of technology, or that they had enough of computers and technology from work. In a previous study, investigating women who chose *not* to participate in screening for colorectal cancer, data indicated that only 1 in 4 actually made use of information presented on the Web. When looking for reasons for this *nonuse*, the only association found was between age and actual Web use. This meant that in older age groups, it was less likely that the women had accessed the information delivered through ICT [30]. The same result, concerning age and ICT, was also indicated in other studies on family members of persons with cancer [31,32]. This might partly reflect the fact that the elderly of today have less experience of using technology than future groups will have and that it is this not only age per se that is hindering the use of ICT. A recent study investigated what might affect the adoption of different technologies (not only ICT) in the elderly, and a factor that seemed important was the *perceived value* [33]. A reflection in relation to this is that, what may be seen as valuable, might also reflect general experience and expectations concerning ICT, and this in turn could be affected by age in today's elderly. Introducing ICT as an intervention aimed at family members might be complex, and *access* may not be the only thing to consider—especially in countries such as Sweden where 94% to 100% of persons aged 16 to 64 years have access to the internet in their homes. This number decreases with older age groups, but still, 86% of persons aged between 65 to 74 years and 68% of persons aged between 75 to 85 years report having internet access in their homes [34].

We found that some family members seemed to be influenced by examples of ICT use, and it might be that, when introduced to ICT, some elderly persons or those with the traits of a *blunter* might need extra support in seeing the usefulness of it and also actually using it. Factors such as age, coping style, perceived value, interest in, and knowledge about ICT could be of importance when developing, introducing, or implementing ICT-based support.

### Limitations

This study was performed in 1 center in 1 country, which may limit transferability to other settings and cultures. However, we believe that some of the experiences may be universal and also apply to caregivers of patients with other chronic conditions. There was a variation in group size between the 8 focus groups. A total of 6 groups had 2 participants each, 1 group 3, and another 8. In the literature [35], it has been discussed that it is the involvement from participants that matters more than the actual group size. Focus group studies entailing only 2 members have been published before [36], and considering the aim and data, we found that group sizes gave relevant data. No new subcategories emerged in the last focus group, which indicated that we had sufficient material to mirror Swedish-born, married/cohabitant caregiver's perceptions of technology from the 8 group discussions.

All researchers, except the main author, were experienced in working with focus group discussions, and there were no significant relationships between researchers and participants.

All but 1 participant was a spouse or partner to the person with HF, and all the participants were of a Swedish background. Most participants were women, and this reflects the population of older family caregivers in particular. The groups reflect different educational background and employment status. It was not possible to recruit participants according to the initial plan, and it might be that carers with other relationships to the ill person and with a more diverse background could have nuanced the results even more. Even so, heterogeneity and homogeneity concerning composition of focus groups has been problematized, and it seems that *commonality* in focus groups is important for capturing the shared experiences and that having too much spread in groups could hinder discussions [37]. We therefore consider the partially homogeneous groups as both a weakness and a strength.

Family members in this study did not always see themselves as caregivers but rather as family. It is possible that family members who lived/supported someone with a more severe HF could more easily have expressed needs in relation to ICT. At the same time, the person with HF defined who to invite to participate in the focus groups, and this strengthens the possibility that our data reflect the general family member population.

### Conclusions

Family members described multiple uses for ICT, and agreed that ICT could provide access to relevant sources of information from which family members potentially could exchange support. ICT was also considered to have its limitations and was out of scope for some but with expected use in future. Even though some family members seemed hesitant about ICT solutions in general, this might not mean they are unreceptive to suggestions about their use in, for example, health care. Thus, a variety of factors should be considered to facilitate future implementations of ICT tools in clinical practice.

## Acknowledgments

The authors are grateful to the family members who chose to take part in the focus groups. Their input has created valuable insights for this study as well as guiding them in developing a support program for informal carers of persons with HF. They would also like to acknowledge the financial support from the Medical Research Council of Southeast Sweden (FORSS-665001), the Swedish National Science Council (VR) K2015-99X -22124-04-4, and Swedish National Science Council/Swedish Research Council for Health, Working Life, and Welfare (VR-FORTE) 2014-4100.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Semistructured interview guide describing opening, introductory, transition, and key questions.

[PDF File (Adobe PDF File), 33KB - jmir_v21i7e13521_app1.pdf ]

## References

1. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 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 J Heart Fail 2016 Dec;18(8):891-975 [FREE Full text] [doi: 10.1002/ejhf.592] [Medline: 27207191]
2. Savarese G, Lund LH. Global public health burden of heart failure. Card Fail Rev 2017 Apr;3(1):7-11 [FREE Full text] [doi: 10.15420/cfr.2016:25:2] [Medline: 28785469]

3. Graven LJ, Martorella G, Gordon G, Grant KJ, Higgins MK. Predictors of depression in outpatients with heart failure: an observational study. Int J Nurs Stud 2017 Apr;69:57-65. [doi: 10.1016/j.ijnurstu.2017.01.014] [Medline: 28182959]
4. Friedmann E, Son H, Thomas SA, Chapa DW, Lee HJ. Poor social support is associated with increases in depression but not anxiety over 2 years in heart failure outpatients. J Cardiovasc Nurs 2014;29(1):20-28 [FREE Full text] [doi: 10.1097/JCN.0b013e318276fa07] [Medline: 23321780]
5. Grant JS, Graven LJ. Problems experienced by informal caregivers of individuals with heart failure: an integrative review. Int J Nurs Stud 2018 Apr;80:41-66. [doi: 10.1016/j.ijnurstu.2017.12.016] [Medline: 29353711]
6. Nicholas DO, Hooker SA, Bekelman D, Ejem D, McGhan G, Kitko L, et al. Family caregiving for persons with heart failure at the intersection of heart failure and palliative care: a state-of-the-science review. Heart Fail Rev 2017 Dec;22(5):543-557 [FREE Full text] [doi: 10.1007/s10741-017-9597-4] [Medline: 28160116]
7. Pressler SJ, Gradus-Pizlo I, Chubinski SD, Smith G, Wheeler S, Sloan R, et al. Family caregivers of patients with heart failure: a longitudinal study. J Cardiovasc Nurs 2013;28(5):417-428. [doi: 10.1097/JCN.0b013e3182563877] [Medline: 22760173]
8. Sullivan BJ, Marcuccilli L, Sloan R, Gradus-Pizlo I, Bakas T, Jung M, et al. Competence, compassion, and care of the self: family caregiving needs and concerns in heart failure. J Cardiovasc Nurs 2016;31(3):209-214 [FREE Full text] [doi: 10.1097/JCN.0000000000000241] [Medline: 25658185]
9. Buck HG, Harkness K, Wion R, Carroll SL, Cosman T, Kaasalainen S, et al. Caregivers' contributions to heart failure self-care: a systematic review. Eur J Cardiovasc Nurs 2015 Feb;14(1):79-89. [doi: 10.1177/1474515113518434] [Medline: 24399843]
10. Kitko LA, Hupcey JE. The work of spousal caregiving of older adults with end-stage heart failure. J Gerontol Nurs 2013 Jul;39(7):40-47. [doi: 10.3928/00989134-20130415-05] [Medline: 23614384]
11. Strömberg A. The situation of caregivers in heart failure and their role in improving patient outcomes. Curr Heart Fail Rep 2013 Sep;10(3):270-275. [doi: 10.1007/s11897-013-0146-8] [Medline: 23857163]
12. Kang X, Li Z, Nolan MT. Informal caregivers' experiences of caring for patients with chronic heart failure: systematic review and metasynthesis of qualitative studies. J Cardiovasc Nurs 2011;26(5):386-394. [doi: 10.1097/JCN.0b013e3182076a69] [Medline: 21263337]
13. Davidson PM, Abernethy AP, Newton PJ, Clark K, Currow DC. The caregiving perspective in heart failure: a population based study. BMC Health Serv Res 2013 Sep 3;13:342 [FREE Full text] [doi: 10.1186/1472-6963-13-342] [Medline: 24004964]
14. Strøm A, Andersen KL, Korneliussen K, Fagermoen MS. Being 'on the alert' and 'a forced volunteer': a qualitative study of the invisible care provided by the next of kin of patients with chronic heart failure. J Multidiscip Healthc 2015;8:271-277 [FREE Full text] [doi: 10.2147/JMDH.S82239] [Medline: 26082643]
15. Blusi M, Kristiansen L, Jong M. Exploring the influence of internet-based caregiver support on experiences of isolation for older spouse caregivers in rural areas: a qualitative interview study. Int J Older People Nurs 2015 Sep;10(3):211-220 [FREE Full text] [doi: 10.1111/opn.12074] [Medline: 25425070]
16. Gusdal AK. Family Caregiving for Persons with Heart Failure: Perspectives of Family Caregivers, Persons With Heart Failure and Registered Nurses. Västerås, Sweden: School of Health, Care and Social Welfare, Mälardalen University; 2017.
17. Adler R, Mehta R. National Alliance for Caregiving. 2014. Catalyzing Technology to Support Family Caregiving URL: https://tinyurl.com/qxym3e8 [accessed 2017-09-27]
18. Chi NC, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2015 Jan;21(1):37-44 [FREE Full text] [doi: 10.1177/1357633X14562734] [Medline: 25475220]
19. Wasilewski MB, Stinson JN, Cameron JI. Web-based health interventions for family caregivers of elderly individuals: a scoping review. Int J Med Inform 2017 Dec;103:109-138. [doi: 10.1016/j.ijmedinf.2017.04.009] [Medline: 28550996]
20. 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]
21. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice, Fourth Edition. Thousand Oaks, California: Sage Publications; 2015.
22. Krueger RA, Casey MA. Focus Groups: A Practical Guide for Applied Research. Thousand Oaks, California: Sage Publications; 2015.
23. Liljeroos M, Agren S, Jaarsma T, Strömberg A. Perceived caring needs in patient-partner dyads affected by heart failure: a qualitative study. J Clin Nurs 2014 Oct;23(19-20):2928-2938. [doi: 10.1111/jocn.12588] [Medline: 24698101]
24. Gusdal AK, Josefsson K, Adolfsson ET, Martin L. Informal caregivers' experiences and needs when caring for a relative with heart failure: an interview study. J Cardiovasc Nurs 2016;31(4):E1-E8. [doi: 10.1097/JCN.0000000000000210] [Medline: 25419945]
25. Washington KT, Meadows SE, Elliott SG, Koopman RJ. Information needs of informal caregivers of older adults with chronic health conditions. Patient Educ Couns 2011 Apr;83(1):37-44. [doi: 10.1016/j.pec.2010.04.017] [Medline: 20452165]
26. Miller SM. Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. J Pers Soc Psychol 1987 Feb;52(2):345-353. [Medline: 3559895]

27. Hickman Jr RL, Daly BJ, Douglas SL, Clochesy JM. Informational coping style and depressive symptoms in family decision makers. Am J Crit Care 2010 Sep;19(5):410-420 [FREE Full text] [doi: 10.4037/ajcc2010354] [Medline: 20810416]
28. Ben-Zur H. Monitoring/blunting and social support: associations with coping and affect. Int J Stress Manag 2002;9(4):357-373 [FREE Full text] [doi: 10.1023/A:1019990518144]
29. Bergström AL, Hanson E. An integrative review of information and communication technology based support interventions for carers of home dwelling older people. Technol Disabil 2018 Feb 5;29(1-2):1-14 [FREE Full text] [doi: 10.3233/TAD-160158] [Medline: 29527109]
30. Fleisher L, Kandadai V, Keenan EM, Miller SM, Devarajan K, Ruth KJ, et al. Build it, and will they come? Unexpected findings from a study on a web-based intervention to improve colorectal cancer screening. J Health Commun 2012;17(1):41-53 [FREE Full text] [doi: 10.1080/10810730.2011.571338] [Medline: 22217118]
31. Lapid MI, Atherton PJ, Clark MM, Kung S, Sloan JA, Rummans TA. Cancer caregiver: perceived benefits of technology. Telemed J E Health 2015 Nov;21(11):893-902 [FREE Full text] [doi: 10.1089/tmj.2014.0117] [Medline: 26075800]
32. Köhle N, Drossaert CH, van Uden-Kraan CF, Schreurs KM, Hagedoorn M, Verdonck-de Leeuw IM, et al. Intent to use a web-based psychological intervention for partners of cancer patients: associated factors and preferences. J Psychosoc Oncol 2018;36(2):203-221. [doi: 10.1080/07347332.2017.1397831] [Medline: 29336704]
33. Berkowsky RW, Sharit J, Czaja SJ. Factors predicting decisions about technology adoption among older adults. Innov Aging 2018 Jan;2(1):igy002 [FREE Full text] [doi: 10.1093/geroni/igy002] [Medline: 30480129]
34. Statistics Database. 2018. [Internet access in Sweden in relation to different age groups] URL: https://tinyurl.com/y56fh8mg [accessed 2019-01-18]
35. Dahlin Ivanoff S. Focus group discussions as a tool for developing a health education programme for elderly persons with visual impairment. Scand J Occup Ther 2002;9(1):3-9. [doi: 10.1080/110381202753505809]
36. Öberg U, Orre CJ, Isaksson U, Schimmer R, Larsson H, Hörnsten A. Swedish primary healthcare nurses' perceptions of using digital eHealth services in support of patient self-management. Scand J Caring Sci 2018;32(2):961-970. [doi: 10.1111/scs.12534] [Medline: 28960451]
37. Ivanoff SD, Hultberg J. Understanding the multiple realities of everyday life: basic assumptions in focus-group methodology. Scand J Occup Ther 2006 Jun;13(2):125-132. [doi: 10.1080/11038120600691082] [Medline: 16856469]

## Abbreviations

**HF:** heart failure
**ICT:** information and communications technology

*Edited by G Eysenbach; submitted 01.02.19; peer-reviewed by A Durante, A Terrill, M Nomali; comments to author 04.04.19; revised version received 12.05.19; accepted 27.05.19; published 16.07.19.*

*Please cite as:*
*Allemann H, Thylén I, Ågren S, Liljeroos M, Strömberg A*
*Perceptions of Information and Communication Technology as Support for Family Members of Persons With Heart Failure: Qualitative Study*
*J Med Internet Res 2019;21(7):e13521*
*URL: http://www.jmir.org/2019/7/e13521/*
*doi:10.2196/13521*
*PMID:31313662*

Original Paper

# Factors Affecting Patients' Use of Electronic Personal Health Records in England: Cross-Sectional Study

Alaa Abd-Alrazaq[1,2], PhD; Bridgette M Bewick[1], PhD; Tracey Farragher[1], PhD; Peter Gardner[3], PhD

[1]Leeds Institute of Health Sciences, School of Medicine, University of Leeds, London, United Kingdom

[2]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[3]School of Psychology, University of Leeds, Leeds, United Kingdom

**Corresponding Author:**
Alaa Abd-Alrazaq, PhD
Division of Information and Computing Technology
College of Science and Engineering
Hamad Bin Khalifa University
Apartment 3
21 Riding House Street
Doha, 00000
Qatar
Phone: 974 55708549
Fax: 974 55708549
Email: aabdalrazaq@hbku.edu.qa

## Abstract

**Background:** Electronic personal health records (ePHRs) are secure Web-based tools that enable individuals to access, manage, and share their medical records. England recently introduced a nationwide ePHR called Patient Online. As with ePHRs in other countries, adoption rates of Patient Online remain low. Understanding factors affecting patients' ePHR use is important to increase adoption rates and improve the implementation success of ePHRs.

**Objective:** This study aimed to examine factors associated with patients' use of ePHRs in England.

**Methods:** The unified theory of acceptance and use of technology was adapted to the use of ePHRs. To empirically examine the adapted model, a cross-sectional survey of a convenience sample was carried out in 4 general practices in West Yorkshire, England. Factors associated with the use of ePHRs were explored using structural equation modeling.

**Results:** Of 800 eligible patients invited to take part in the survey, 624 (78.0%) returned a valid questionnaire. Behavioral intention (BI) was significantly influenced by performance expectancy (PE; beta=.57, *P*<.001), effort expectancy (EE; beta=.16, *P*<.001), and perceived privacy and security (PPS; beta=.24, *P*<.001). The path from social influence to BI was not significant (beta=.03, *P*=.18). Facilitating conditions (FC) and BI significantly influenced use behavior (UB; beta=.25, *P*<.001 and beta=.53, *P*<.001, respectively). PE significantly mediated the effect of EE and PPS on BI (beta=.19, *P*<.001 and beta=.28, *P*=.001, respectively). Age significantly moderated 3 paths: PE→BI, EE→BI, and FC→UB. Sex significantly moderated only the relationship between PE and BI. A total of 2 paths were significantly moderated by education and internet access: EE→BI and FC→UB. Income moderated the relationship between FC and UB. The adapted model accounted for 51% of the variance in PE, 76% of the variance in BI, and 48% of the variance in UB.

**Conclusions:** This study identified the main factors that affect patients' use of ePHRs in England, which should be taken into account for the successful implementation of these systems. For example, developers of ePHRs should involve patients in the process of designing the system to consider functions and features that fit patients' preferences and skills to ensure systems are useful and easy to use. The proposed model accounted for 48% of the variance in UB, indicating the existence of other, as yet unidentified, factors that influence the adoption of ePHRs. Future studies should confirm the effect of the factors included in this model and identify additional factors.

***(J Med Internet Res 2019;21(7):e12373)*** doi:10.2196/12373

**KEYWORDS**

health records, personal; patient portal; electronic personal health records; technology acceptance; technology adoption; intention; unified theory of acceptance and use of technology; structural equation modelling

## Introduction

### Background

Electronic personal health records (ePHRs) refer to secure Web-based tools that enable individuals to access and manage their medical records and share them with trusted others [1]. More advanced ePHRs provide additional functionalities, such as scheduling appointments, requesting prescription refills, messaging providers, requesting referrals, and educational tools [2-4]. Benefits of using ePHRs include the following: enhancing patient empowerment [5,6], improving patient self-management and medication adherence [7,8], enhancing the relationships and communications between patients and health care providers [9,10], enabling patients to easily access health services [11,12], avoiding duplicated tests [9,11], and reducing adverse drug interactions and allergies [9,11,13].

In 2015, the National Health Service in England launched a program called Patient Online, which requires general practices (GPs) to provide patients with Web-based services, such as booking appointments, requesting prescription refills, and viewing summary information from GP records [14,15]. GPs use one of the following systems to provide their patients with the abovementioned services: SystemOnline, Patient Access, Patient Services, The Waiting Room, Engage Consult, and Evergreen Life or i-Patient [14].

### Research Problem and Aim

The overall adoption rate of Patient Online was 18.9% in April 2017 and reached 24.4% in April 2018 [16], and so adoption remains low. Identifying and understanding factors that affect patients' use of ePHRs is crucial to develop interventions to increase patients' adoption and improve the implementation success of ePHRs [17-22]. According to a systematic review conducted by Abd-alrazaq and colleagues [23], there are no published studies on factors affecting patients' use of ePHRs in England. Although many studies have been conducted in other countries, they have several shortcomings, namely, (1) few studies were theory-based research [21,24-27], (2) many studies focused on factors that affect patients' intention to use ePHRs instead of actual use [28-32], (3) many studies have assessed the factors that affect self-reported use rather than actual use [27,32-35], (4) almost all studies examined independent and dependent variables at one point in time using the same data collection instrument, so being at risk of common method bias [25,32,36], and (5) almost all studies did not differentiate between factors affecting initial use and continuing use of ePHRs.

This study aimed to examine factors associated with patients' adoption of ePHRs (Patient Online) in England. As 76% of patients in England have never used Patient Online [16], the study focused on factors associated with patients' initial use of ePHRs. Therefore, it was more appropriate to investigate the factors that make nonusers become users (ie, initial use stage).

## Methods

### Theoretical Foundation

In total, 12 theories and models originated from various disciplines, such as psychology, sociology, and information systems, were reviewed to select the appropriate one for our study. Selection of the appropriate theory was based on predefined 6 criteria. Although 2 criteria were related to the applicability of the theory on the phenomena of interest (ie, population and type of behavior), the remaining 4 were related to goodness of the theory (ie, logical consistency, explanatory power, falsifiability, and parsimony). The unified theory of acceptance and use of technology (UTAUT) was the only theory that met all those criteria. Therefore, this study chose UTAUT as a theoretical lens to examine factors associated with patients' use of ePHRs. More details about how the theories met or did not meet each criterion are explained in Multimedia Appendix 1.

According to UTAUT, behavioral intention (BI) is affected directly by performance expectancy (PE), effort expectancy (EE), and social influence (SI) [37]. Both BI and facilitating conditions (FC) are hypothesized to affect use behavior (UB) directly [37]. UTAUT also proposes that most of these relationships are moderated by age, sex, experience, and voluntariness [37].

In this study, the adoption of ePHRs is not compulsory. The UTAUT construct of voluntariness is only applicable in nonvoluntary contexts [38]. Thus, for this study, the moderator *voluntariness* was dropped from the model. This study focused on the factors that explained how nonusers become users of ePHRs (ie, preusage stage); the sample comprised only nonusers of ePHRs (ie, having no experience). For that reason, the moderator *experience* was also removed from the model.

A review of the literature identified a consensus on the influential effect of the following factors on ePHRs adoption: PPS [26,39-48], internet access [11,28,39,49-53], income [26,28,39,49,51,54-58], and education level [26,28,39,44,49,51,56,59-63]. These 4 factors were not part of UTAUT but were included in our adapted model to make it more appropriate for the context of ePHRs adoption. Although PPS was proposed as an independent variable, the remaining 3 factors were hypothesized as a moderator. The research hypotheses and the proposed model are presented in Table 1 and Figure 1, respectively. Multimedia Appendix 2 shows the conceptual definitions of the constructs in the proposed model. Multimedia Appendix 3 shows the theoretical foundations for the new proposed relationships that were added to the UTAUT model.

**Table 1.** The research hypotheses.

| H[a] number | Hypothesis |
|---|---|
| H1 | PE[b] positively influences patients' intention to use Patient Online. |
| H2 | Age, sex, education, and income moderate the positive relationship between PE and patients' intention to use Patient Online, such that the influence is stronger for younger males with lower level of education and higher income. |
| H3 | EE[c] positively influences patients' intention to use Patient Online. |
| H4 | PE positively mediates the positive relationship between EE and BI[d]. |
| H5 | Age, sex, education, income, and internet access moderate the positive relationship between EE and patients' intention to use Patient Online, such that the influence is stronger for older females with lower level of education and income and without internet access. |
| H6 | SI[e] positively influences patients' intention to use Patient Online. |
| H7 | Age and sex moderate the positive relationship between SI and patients' intention to use Patient Online, such that the influence is stronger for older females. |
| H8 | PPS[f] positively influences patients' intention to use Patient Online. |
| H9 | PE positively mediates the positive relationship between PPS and BI. |
| H10 | Age, sex, education, and income moderate the positive relationship between PPS and patients' intention to use Patient Online, such that the influence is stronger for older females with higher level of education and lower income. |
| H11 | FC[g] positively influences patients' use of Patient Online. |
| H12 | Age, sex, education, income, and internet access moderate the positive relationship between FC and UB[h], such that the influence is stronger for older females with a lower level of education and income and without internet access. |
| H13 | BI positively influences patients' use of Patient Online. |

[a]H: hypothesis.

[b]PE: performance expectancy.

[c]EE: effort expectancy.

[d]BI: behavioral intention.

[e]SI: social influence.

[f]PPS: perceived privacy and security.

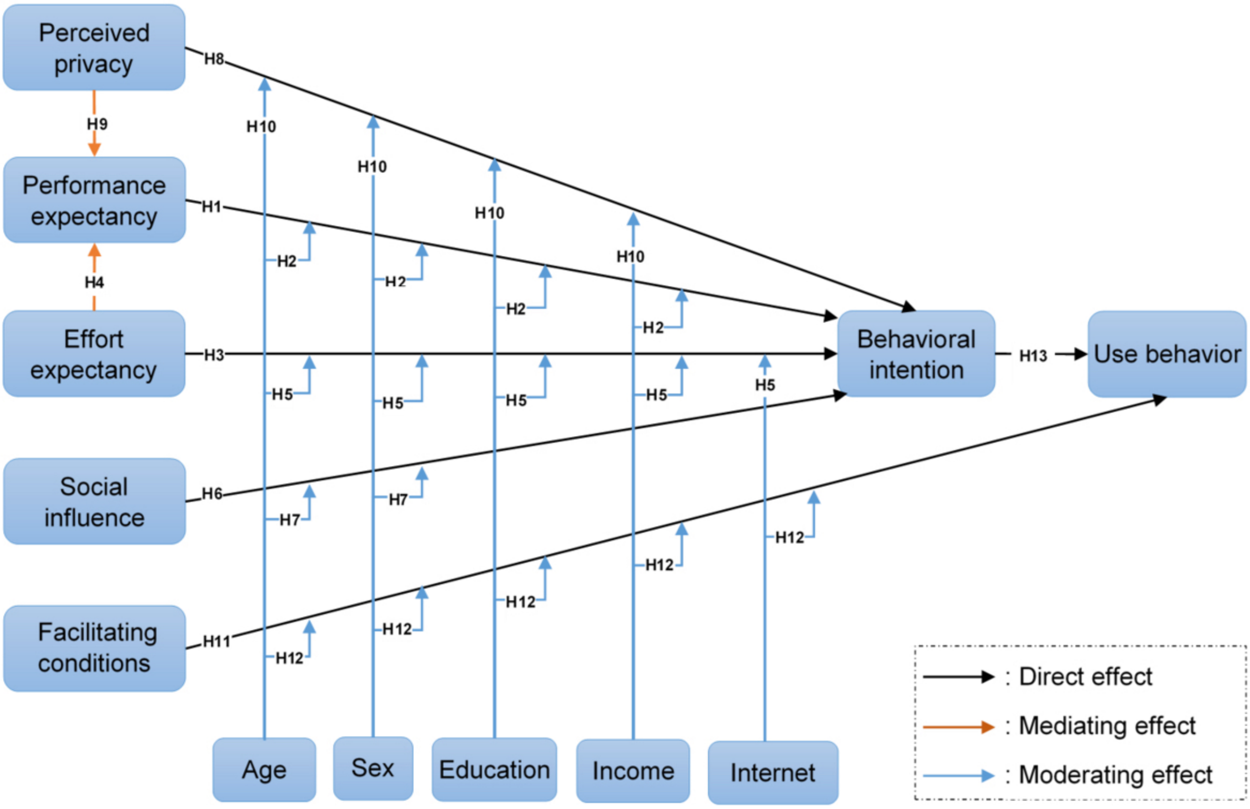
[g]FC: facilitating conditions.

[h]UB: use behavior.

## Study Design and Setting

The proposed model was examined empirically using data from a cross-sectional survey. The survey was conducted at 4 West Yorkshire (England) GPs, 3 practices in Bradford and 1 in Leeds. More details about these practices are shown in Multimedia Appendix 4. Health Research Authority approval for this study was granted before starting data collection (REC reference: 17/SC/0323).

**Figure 1.** The proposed model.




### Measurement

Self-administrated questionnaires were used to measure all variables proposed in the model except UB. UB was measured objectively using system logs that recorded the use of PatientOnline. Questionnaires included 29 well-validated items adopted from previous studies (Multimedia Appendix 5). An introduction about Patient Online was included at the top of the questionnaire to ensure all participants had the knowledge necessary to answer questions about Patient Online. The questionnaire was validated by sending it to a panel of experts to assess the face validity and content validity of the questions. After modifying the questionnaire according to experts' recommendations, it was pilot tested by sending it via email to 37 patients (members of patient and carer community) who were asked to fill in the questionnaire and answer questions regarding clarity or ambiguity of questions, clarity of instructions to answer questions, difficulty to answer questions, time needed to complete the questionnaire, clarity and attractiveness of the layout, missing of important topics, and sequence of questions. A few issues were reported by experts and patients, and the questionnaire was modified accordingly (Multimedia Appendix 6). System log data of the number of times that each participant logged into the system during 6 months after completing the questionnaire were the objective measure of use.

### Recruitment

We recruited a convenience sample of patients from August 21, 2017, to September 26, 2017. Patients were eligible to participate if they (1) lived in England and were registered at 1 of the 4 GPs, (2) were aged 18 years or older, and (3) had not used Patient Online before (nonusers). The researcher distributed the questionnaire to eligible participants visiting 1 of the 4 GPs. After 6 months from the completion of the questionnaire, data from the system log were extracted to ascertain participants' use of Patient Online.

### Statistical Analysis

Structural equation modeling (SEM) was used to test the theoretical model and hypotheses. Specifically, the measurement model was examined in terms of 3 aspects: model fit, construct reliability, and construct validity [64,65]. After ensuring the validity of the measurement model, the structural model was tested in terms of 3 aspects: model fit, predictive power, and strength of relationships [65-67]. The strength of relationships was examined using different methods depending on the type of the proposed effect. Specifically, direct effects were assessed by checking path coefficients [68]. Mediating effects were examined by assessing the indirect effect using bootstrapping. The moderating effect for the metric moderator (ie, age) was examined using the interaction effect method [64,69]. The moderating effects for nonmetric moderators were tested using multigroup SEM [64,69,70]. Analysis of moment structures (version 24; IBM SPSS) software was used for conducting all abovementioned analyses.

## Results

### Participants' Characteristics

Of the 800 eligible patients invited to take part in the survey, 624 participants returned a completed questionnaire giving a response rate of 78%. The mean age of participants was 44.2

years. The majority of participants were white (79.8%, 498/624) and had internet access (84.6%, 528/624; Table 2). Differences between participants and nonparticipants in terms of age, sex, and ethnicity were not significant (*P*=.21, *P*=.06, and *P*=.64, respectively). It was, therefore, concluded that the risk of nonresponse bias was minimal.

**Table 2.** Participants' characteristics (n=624).

| Variables | Respondents, n (%) |
|---|---|
| **Age (years)**[a] | |
| 18-24 | 107 (17.1) |
| 25-34 | 148 (23.7) |
| 35-44 | 116 (18.6) |
| 45-54 | 98 (15.7) |
| 55-64 | 65 (10.4) |
| 65-74 | 46 (7.4) |
| 75 and older | 44 (7.1) |
| **Sex** | |
| Male | 293 (46.9) |
| Female | 331 (53.1) |
| **Ethnicity** | |
| White | 498 (79.8) |
| Asian | 73 (11.7) |
| Black | 20 (3.2) |
| Mixed | 26 (4.1) |
| Others | 7 (1.2) |
| **Income (£)** | |
| <20,000 | 284 (45.5) |
| 20,000-29,999 | 80 (12.8) |
| 30,000-39,999 | 65 (10.4) |
| 40,000-49,999 | 43 (7.0) |
| 50,000-59,999 | 26 (4.2) |
| ≥60,000 | 12 (1.9) |
| Prefer not to say | 114 (18.2) |
| **Education** | |
| Up to secondary school | 69 (11.1) |
| Secondary school | 147 (23.6) |
| College/diploma | 165 (26.4) |
| Bachelor's degree | 174 (27.9) |
| Master's degree | 47 (7.5) |
| Doctoral degree | 22 (3.5) |
| **Internet access** | |
| Yes | 528 (84.6) |
| No | 96 (15.4) |

[a]Mean 44.2 (SD 18.9).

## Measurement Model

### Model Fit

All fit indices indicated a good fit of the initial model except the following 3 indices: goodness-of-fit index (GFI, 0.923), root mean square error of approximation (RMSEA, 0.053), and standardized root mean square residual (SRMR, 0.057; Table 3). The following 4 items were identified as a source of the poor fit of the measurement model as their factor loading was less than 0.70: FC 4, FC 5, PPS 3, and PPS 5. After deleting these 4 items from the model, all fit indices of the modified model improved and existed within their acceptable levels, indicating a good fit (Table 3).

**Table 3.** Results of fit indices of the initial and modified measurement model.

| Fit indices | Cutoff point | Initial measurement model | Modified measurement model |
|---|---|---|---|
| Relative chi-square (*df*) | 1-3 | 2.8 (215) | 1.4 (137) |
| GFI[a] | ≥0.95 | 0.923 | 0.969 |
| AGFI[b] | ≥0.90 | 0.902 | 0.957 |
| RMSEA[c] | <0.05 | 0.053 | 0.026 |
| PCLOSE[d] | ≥0.05 | 0.194 | 1.000 |
| SRMR[e] | ≤0.05 | 0.057 | 0.017 |
| NFI[f] | ≥0.95 | 0.964 | 0.988 |
| CFI[g] | ≥0.95 | 0.977 | 0.995 |
| TLI[h] | ≥0.95 | 0.972 | 0.996 |

[a]GFI: goodness-of-fit index.
[b]AGFI: adjusted goodness-of-fit index.
[c]RMSEA: root mean square error of approximation.
[d]PCLOSE: p of close fit.
[e]SRMR: standardized root mean square residual.
[f]NFI: normed fit index.
[g]CFI: comparative fit index.
[h]TLI: Tucker-Lewis index.

### Construct Reliability

Results for the modified model of Cronbach alpha, composite reliability, and average variance extracted (AVE) for each construct were within their cutoff of ≥.70, ≥0.70, and ≥0.50, respectively (Multimedia Appendix 7). This indicates that the measurement items are consistent and reproducible in measuring what it is assumed to measure.

### Construct Validity

The values of factor loading and AVE for all items considerably exceeded the thresholds of 0.70 and 0.50, respectively (Multimedia Appendix 8). These results indicate that items had good convergent validity. Similarly, items showed good discriminant validity according to 3 measures. Specifically, intercorrelation coefficients are located within the acceptable ranges (<0.85; Multimedia Appendix 9). With regard to the second measure, each value of square root of AVE for a construct (values on the diagonal) was higher than all intercorrelation coefficients between that construct and each other construct (off-diagonal values; Multimedia Appendix 9). With regard to the third measure, the loading of each item on its construct was higher than cross-loadings in rows and columns (Multimedia Appendix 10).

## Structural Model

### Model Fit and Predictive Power

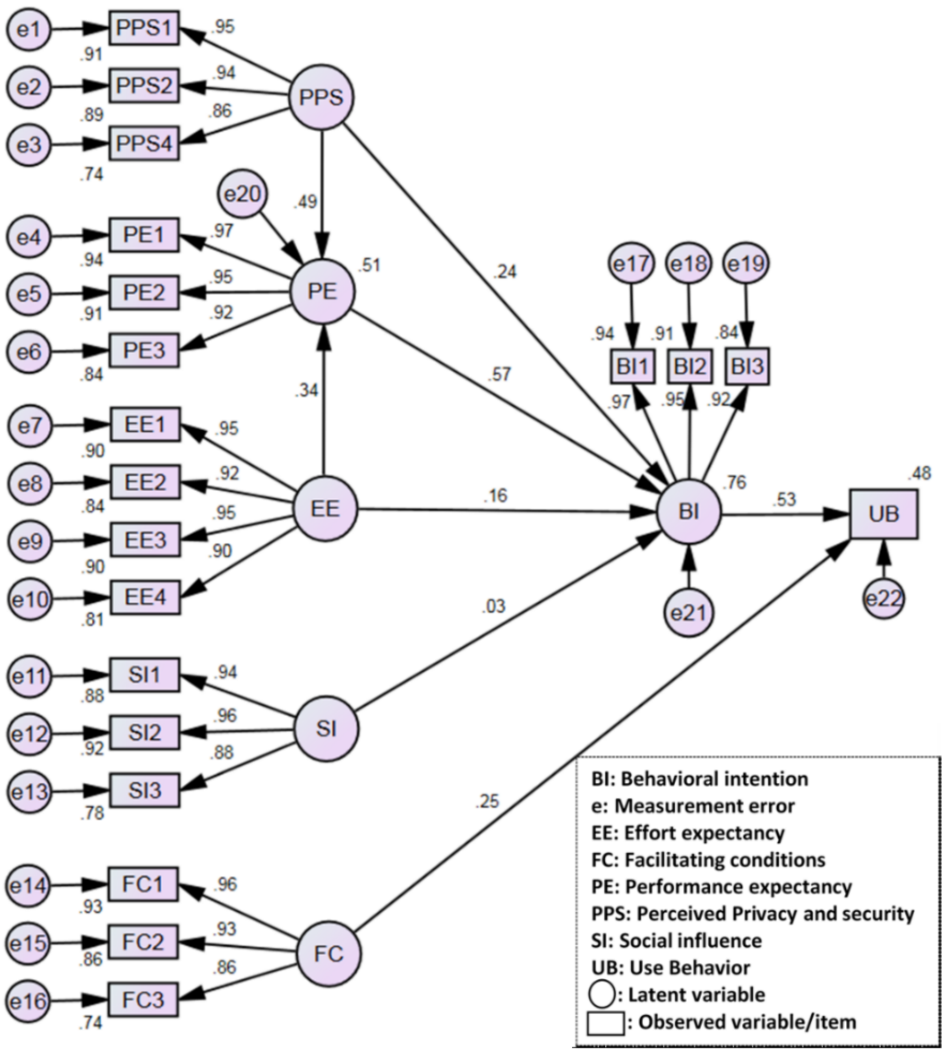
All fit indices of the structural model indices were within their cutoff levels, indicating a good model fit (Table 4). The structural model accounted for 51% of the variance in PE, 76% of the variance in BI, and 48% of the variance in UB (Figure 2).

**Table 4.** Results of fit indices of the structural model.

| Fit indices | Cutoff point | Fitness of the structural model |
|---|---|---|
| Relative chi-square (*df*) | 1-3 | 1.6 (157) |
| GFI[a] | ≥0.95 | 0.962 |
| AGFI[b] | ≥0.90 | 0.949 |
| RMSEA[c] | <0.05 | 0.032 |
| PCLOSE[d] | ≥0.05 | 1.000 |
| SRMR[e] | ≤0.05 | 0.036 |
| NFI[f] | ≥0.95 | 0.984 |
| CFI[g] | ≥0.95 | 0.993 |
| TLI[h] | ≥0.95 | 0.992 |

[a]GFI: goodness-of-fit index.
[b]AGFI: adjusted goodness-of-fit index.
[c]RMSEA: root mean square error of approximation.
[d]PCLOSE: p of close fit.
[e]SRMR: standardized root mean square residual.
[f]NFI: normed fit index.
[g]CFI: comparative fit index.
[h]TLI: Tucker-Lewis index.

**Figure 2.** Structural model estimates.




### *Strength of Relationships*

Of the direct effects, BI was associated with PE (beta=.57), EE (beta=.16), and PPS (beta=.24; Table 5). The path from SI to BI was not significant (beta=.03, *P*=.18). Both FC and BI were significantly associated with UB (beta=.25 and beta=.53, respectively). Therefore, the following hypotheses were supported: H1, H3, H8, H11, and H13 (Table 5).

With regard to mediating effects, results of bootstrapping indicate that PE mediated significantly the effect of EE and PPS on BI (beta=.20 and beta=.28, respectively; Table 6). Accordingly, H4 and H9 were supported in this research.

With regard to moderating effects, age moderated significantly 3 paths: PE→BI (beta=−.10), EE→BI (beta=.06), and FC→UB (beta=.16; Table 7). Sex moderated significantly only the relationship between PE and BI (*P*=.009; Table 8). In relation to moderating effect of education, the relationship between FC and UB was statistically stronger for *secondary school or lower* group than college group (beta=.39 vs beta=.30, *P*=.003; Table 9) and than *bachelor or higher* group (beta=.39 vs beta=.21; Table 10). The path from EE to BI was statistically weaker for *bachelor or higher* group than *secondary school or lower* group (beta=.01 vs beta=.14; Table 10) and than college group (beta=.01 vs beta=.13; Table 11). As shown in Table 12-14, the relationship between FC and UB was statistically stronger for patients with low income (beta=.43) than patients with moderate or high income (beta=.25 and beta=.10, respectively). Internet access moderated significantly 2 paths EE→BI (*P*=.01) and FC→UB (*P*<.001; Table 15). Accordingly, H10 was rejected, whereas the following hypotheses were partially supported: H2, H5, H7, and H12.

**Table 5.** Results of direct effects.

| H[a] | Path | SE (beta) | 95% CI | *P* value | Supported? |
|---|---|---|---|---|---|
| H1 | PE[b]→BI[c] | .57 | 0.51 to 0.64 | <.001 | Yes |
| H3 | EE[d]→BI | .16 | 0.11 to 0.21 | <.001 | Yes |
| H6 | SI[e]→BI | .03 | −0.03 to 0.10 | .18 | No |
| H8 | PPS[f]→BI | .24 | 0.18 to 0.29 | <.001 | Yes |
| H11 | FC[g]→UB[h] | .25 | 0.20 to 0.30 | <.001 | Yes |
| H13 | BI →UB | .53 | 0.48 to 0.58 | <.001 | Yes |

[a]H: hypothesis.
[b]PE: performance expectancy.
[c]BI: behavioral intention.
[d]EE: effort expectancy.
[e]SI: social influence.
[f]PPS: perceived privacy and security.
[g]FC: facilitating conditions.
[h]UB: use behavior.

**Table 6.** Results of mediating effects.

| H[a] | Indirect effect | Estimate (beta) | 95% CI | *P* value | Supported? |
|---|---|---|---|---|---|
| H4 | EE[b]→PE[c]→BI[d] | .20 | 0.15-0.25 | <.001 | Yes |
| H9 | PPS[e]→PE→BI | .28 | 0.23-0.33 | <.001 | Yes |

[a]H: hypothesis.
[b]EE: effort expectancy.
[c]PE: performance expectancy.
[d]BI: behavioral intention.
[e]PPS: perceived privacy and security.

**Table 7.** Results of moderating effect of age.

| Interaction effect | Standardized estimate (beta) | *P* value |
|---|---|---|
| PE[a]×Age→BI[b] | −.10 | <.001 |
| EE[c]×Age→BI | .06 | .03 |
| SI[d]×Age→BI | .01 | .06 |
| PPS[e]×Age→BI | −.03 | .22 |
| FC[f]×Age→UB[g] | .16 | <.001 |

[a]PE: performance expectancy.
[b]BI: behavioral intention.
[c]EE: effort expectancy.
[d]SI: social influence.
[e]PPS: perceived privacy and security.
[f]FC: facilitating conditions.
[g]UB: use behavior.

**Table 8.** Results of moderating effect of sex.

| Hypothesized path | SE (beta) | | *P* value | | Chi-square difference test, *P* value |
|---|---|---|---|---|---|
| | Male | Female | Male | Female | |
| PE[a]→BI[b] | .59 | .51 | <.001 | <.001 | .01 |
| EE[c]→BI | .17 | .19 | <.001 | <.001 | .32 |
| SI[d]→BI | −.03 | .06 | .53 | .06 | .07 |
| PPS[e]→UB | .27 | .20 | <.001 | <.001 | .65 |
| FC[f]→UB[g] | .35 | .28 | <.001 | <.001 | .32 |

[a]PE: performance expectancy.
[b]BI: behavioral intention.
[c]EE: effort expectancy.
[d]SI: social influence.
[e]PPS: perceived privacy and security.
[f]FC: facilitating conditions.
[g]UB: use behavior.

**Table 9.** Results of moderating effect of education level (secondary school versus college/diploma).

| Hypothesized path | Secondary school or lower | | College/diploma | | Chi-square difference test, *P* value |
|---|---|---|---|---|---|
| | SE (beta) | *P* value | SE (beta) | *P* value | |
| PE[a]→BI[b] | .57 | <.001 | .62 | <.001 | .38 |
| EE[c]→BI | .14 | .02 | .13 | .003 | .38 |
| PPS[d]→BI | .17 | .005 | .29 | <.001 | .22 |
| FC[e]→UB[f] | .39 | <.001 | .30 | <.001 | .003 |

[a]PE: performance expectancy.
[b]BI: behavioral intention.
[c]EE: effort expectancy.
[d]PPS: perceived privacy and security.
[e]FC: facilitating conditions.
[f]UB: use behavior.

**Table 10.** Results of moderating effect of education level (secondary school versus bachelor or higher).

| Hypothesized path | Secondary school or lower | | Bachelor or higher | | Chi-square difference test, *P* value |
|---|---|---|---|---|---|
| | SE (beta) | *P* value | SE (beta) | *P* value | |
| PE[a]→BI[b] | .57 | <.001 | .57 | <.001 | .50 |
| EE[c]→BI | .14 | .02 | .01 | .16 | .03 |
| PPS[d]→BI | .17 | .005 | .24 | <.001 | .14 |
| FC[e]→UB[f] | .39 | <.001 | .21 | <.001 | .02 |

[a]PE: performance expectancy.
[b]BI: behavioral intention.
[c]EE: effort expectancy.
[d]PPS: perceived privacy and security.
[e]FC: facilitating conditions.
[f]UB: use behavior.

**Table 11.** Results of moderating effect of education level (college/diploma versus bachelor or higher).

| Hypothesized path | College/diploma | | Bachelor or higher | | Chi-square difference test, *P* value |
|---|---|---|---|---|---|
| | SE (beta) | *P* value | SE (beta) | *P* value | |
| PE[a]→BI[b] | .62 | <.001 | .57 | <.001 | .08 |
| EE[c]→BI | .13 | .003 | .01 | .16 | .008 |
| PPS[d]→BI | .29 | <.001 | .24 | <.001 | .54 |
| FC[e]→UB[f] | .30 | <.001 | .21 | <.001 | .23 |

[a]PE: performance expectancy.
[b]BI: behavioral intention.
[c]EE: effort expectancy.
[d]PPS: perceived privacy and security.
[e]FC: facilitating conditions.
[f]UB: use behavior.

**Table 12.** Results of moderating effect of income (low income versus middle income).

| Hypothesized path | Low income[a] | | Middle income[b] | | Chi-square difference test, *P* value |
|---|---|---|---|---|---|
| | SE (beta) | *P* value | SE (beta) | *P* value | |
| PE[c]→BI[d] | .54 | <.001 | .52 | <.001 | .43 |
| EE[e]→BI | .14 | <.001 | .22 | <.001 | .39 |
| PPS[f]→BI | .26 | <.001 | .28 | <.001 | .99 |
| FC[g]→UB[h] | .43 | <.001 | .25 | <.001 | .048 |

[a]Low income: >£20,000.
[b]Medium income: £20,000-39,999.
[c]PE: performance expectancy.
[d]BI: behavioral intention.
[e]EE: effort expectancy.
[f]PPS: perceived privacy and security.
[g]FC: facilitating conditions.
[h]UB: use behavior.

**Table 13.** Results of moderating effect of income (low income versus high income).

| Hypothesized path | Low income[a] | | High income[b] | | Chi-square difference test, *P* value |
|---|---|---|---|---|---|
| | SE (beta) | *P* value | SE (beta) | *P* value | |
| PE[c]→BI[d] | .54 | <.001 | .68 | <.001 | .09 |
| EE[e]→BI | .14 | <.001 | .12 | .048 | .67 |
| PPS[f]→BI | .26 | <.001 | .25 | <.001 | .87 |
| FC[g]→UB[h] | .43 | <.001 | .10 | .03 | .03 |

[a]Low income: >£20,000.
[b]High income: ≥£40,000.
[c]PE: performance expectancy.
[d]BI: behavioral intention.
[e]EE: effort expectancy.
[f]PPS: perceived privacy and security.
[g]FC: facilitating conditions.
[h]UB: use behavior.

**Table 14.** Results of moderating effect of income (middle income versus high income).

| Hypothesized path | Middle income[a] | | High income[b] | | Chi-square difference test, *P* value |
|---|---|---|---|---|---|
| | SE (beta) | *P* value | SE (beta) | *P* value | |
| PE[c]→BI[d] | .52 | <.001 | .68 | <.001 | .06 |
| EE[e]→BI | .22 | <.001 | .12 | .048 | .27 |
| PPS[f]→BI | .28 | <.001 | .25 | <.001 | .88 |
| FC[g]→UB[h] | .25 | <.001 | .10 | .03 | .02 |

[a]Medium income: £20,000-39,999.
[b]High income: ≥£40,000.
[c]PE: performance expectancy.
[d]BI: behavioral intention.
[e]EE: effort expectancy.
[f]PPS: perceived privacy and security.
[g]FC: facilitating conditions.
[h]UB: use behavior.

**Table 15.** Results of the moderating effect of internet access.

| Hypothesized path | SE (beta) | | *P* value | | Chi-square difference test, *P* value |
|---|---|---|---|---|---|
| | Internet access | No internet access | Internet access | No internet access | |
| EE[a]→BI[b] | .12 | .28 | <.001 | .005 | .01 |
| FC[c]→UB[d] | .18 | .44 | <.001 | <.001 | <.001 |

[a]EE: effort expectancy.
[b]BI: behavioral intention.
[c]FC: facilitating conditions.
[d]UB: use behavior.

## Discussion

### Principal Findings

This study found that PE was positively associated with BI. This means that patients are more likely to intend to use Patient Online when they perceive it to be very useful and advantageous. This relationship is consistent with other studies investigating the uptake of ePHRs [18,26,71,72-75]. Our results suggest that this relationship is stronger for younger males, meaning that younger males who perceive the system as more useful are more likely to intend to use it. This study is one of the very few studies that successfully assessed the moderating effect of age and sex in explaining the use of ePHRs. A similar moderating effect has been demonstrated for the use of consumer health information technologies (CHITs) [76,77].

These results showed that EE was positively associated with BI, that is, patients are more likely to intend to use Patient Online when they perceive it as an easy-to-use system. This finding is consistent with studies investigating the use of ePHRs outside of England [26,72,78-80]. This study showed that the relationship between EE and BI was mediated by PE. So, patients who perceive Patient Online as easy to use are more likely to perceive it as a useful system, thereby, they are more likely to intend to use it. This finding is in line with findings of 2 CHIT studies [22,81]. Furthermore, our results showed that the relationship between EE and BI was stronger among older patients with lower level of education and without internet access. The moderating effect of age has also been found in studies investigating the use of CHITs [76]. Ours is the first study to examine the moderating effect of education and internet access to explain the use of ePHRs or CHITs.

We found that SI and BI were not statistically associated. This means that opinions and beliefs of people who are important to the patient do not affect their intention to utilize ePHRs. This nonsignificant relationship could be attributed to the use of Patient Online being voluntary. The literature suggests that the effect of SI is significant only in contexts where using the technology is mandatory [37,82-84]. The presence of PE in a model may weaken the direct effect of SI on BI [85,86] as SI affects BI indirectly through PE [22,84]. The nonsignificant effect of SI may also be attributed to the fact that the questionnaire measures perceptions of SI rather than actual SI.

This study demonstrated that PPS was positively associated with BI, that is, patients are more likely to intend to use Patient Online when they perceive that it is secure and will maintain their privacy. This relationship is documented elsewhere in the literature [30,45,46,48,87,88]. This study showed that the relationship between PPS and BI was mediated by PE, that is, patients who perceive Patient Online to be secure and able to maintain their privacy are more likely to perceive it as a useful system and, therefore, are more likely to intend to use it. Although several studies in the context of ePHRs and CHITs examined the direct influence of PPS on PE and on BI, no previous study has tested the indirect effect of PPS on BI through PE.

The statistical analysis showed that FC was positively associated with UB. This means that patients are more likely to use Patient Online when they feel that they have the resources and knowledge enough to use it. This effect of FC was supported by several studies in the context of ePHRs [43,44,89]. In our study, this relationship was stronger for older patients with lower level of education and income and without internet access. In other words, these groups of people tend to place more importance on the presence of sufficient support and resources to use Patient Online. Although the moderating effect of age was supported in 1 CHITs study [76], this is the first study to investigate the moderating effects of education, income, and internet access in the context of ePHRs.

This study showed that BI positively associated with UB, that is, patients are more likely to use Patient Online when they intend to use it. This finding is consistent with findings of several studies in the context of ePHRs and CHITs [22,80,90-93].

Overall, the model accounted for 48% of the variance in UB. This moderate predictive power of the model indicates that there are other factors yet to be identified, which would account for the unexplained variance. Although the predictive power of the proposed model is comparable with the predictive power of the original UTAUT model (48%), it is higher than the predictive power of models proposed by other studies in the context of ePHRs: Hsieh [90] (42.7%) and Tavares and Oliveira [80] (26.8%).

### Theoretical and Practical Contributions

This is the first theory-based study to examine factors associated with patients' use of ePHRs in England. Very few studies have utilized theories or models to understand the factors that impact patients' use of ePHRs [21,24-27]. Furthermore, UTAUT was not employed in those few studies. Accordingly, this study contributes to the ePHR literature by adopting and testing UTAUT in the context of ePHRs, which can be used by the future studies in the context of ePHRs and CHITs.

This research and a study conducted by Tavares and Oliveira [80] are the only studies in the area of ePHRs that included both BI and UB in 1 model, and this is the best practice to study technology adoption [20,25]. Furthermore, our study contributes to the existing ePHR literature by being the first theory-based study to measure the UB objectively. In addition, this study is one of the few theory-based studies in the context of ePHRs that endeavored to minimize the common method bias by ensuring a gap in time between the main dependent variable (ie, use of Patient Online) and other variables. This research is the first study to shed light on the important role of moderators and mediators that explain the use of ePHRs, and this extends our understanding of factors that affect the adoption [27].

With respect to practical contributions, we have identified that PE and EE play a crucial role in forming patients' intention to use Patient Online. Accordingly, developers should involve patients in the process of designing the system to consider functions and features that fit patients' preferences and skills. Developers should pilot test the system with potential users before implementation [80,94]. As PPS is an influential predictor, developers should keep patient records as private as possible by protecting the platforms using security measures, such as strong firewalls, complex and long passwords, regular security reviews, and regular website updates.

To ensure that patients perceive the system as useful, easy to use, and secure, marketers should conduct promotional campaigns about functions and features of the system, its advantages, its ease of use, availability of different sources to support the use of the system, the security measures, the laws and regulations protecting patient privacy, and how patients can use it safely. As face-to-face communication may be one of the most effective channels in marketing to persuade potential adopters to adopt an innovation [95-97], physicians, nurses, and receptionists can play an important role in improving the publicity of Patient Online by informing patients about it in their communications. Marketers should focus more on younger males when conducting promotional campaigns regarding the benefits of the system, whereas they should concentrate more on older and less educated patients without internet access when initiating advertising campaigns regarding the ease of use of the system.

Patients who believed that organizational and technical infrastructure existed to support the use of Patient Online were more likely to use it. Therefore, to raise awareness of the infrastructure available, GPs could provide patients with manuals, Web-based assistance, technical support, and practical training sessions. This strategy is likely to be most effective with older patients and those with lower level of education and income.

Allowing patients to try a beta version of ePHRs could create a positive personal experience that may enhance their perceptions of usefulness and the ease of use of the system [26,75,78]. Thus, GPs should assist patients in using a beta version of Patient Online through a computer in a waiting room.

### Research Limitations

This study has limitations that need to be considered. Data were collected from 4 GPs, all implementing the same ePHRs (ie, SystemOnline); therefore, the findings of this study may not be applicable to other ePHRs (ie, Patient Access and Patient Services). However, the findings may still be generalizable to other systems because all the systems mentioned provide the same services to the patients, and all participants had not used any of them before. Therefore, they would be unlikely to have different perceptions about the different systems.

This study focused on assessing factors that affect patients' initial use of ePHRs; therefore, the findings are not generalizable to the context of continuing use. This research focuses on initial use of ePHRs because Patient Online is still a new system in England and has a low adoption rate; therefore, it is better to focus on the initial use in this period. Furthermore, as this study is cross-sectional, the associations identified with patients' initial use of ePHRs do not imply causality, and so further longitudinal research is required.

This research is subject to a sampling bias because of using convenience sampling technique to recruit the participants [36,98]. This study found that there was no statistically significant difference between the participants and the nonparticipants in terms of age, sex, and ethnicity. Consequently, it can be said that the findings are generalizable to practices similar to the 4 practices in this study.

It might have been appropriate to control for the effects from practices within the SEM. However, we found no differences in the demographics between the practices, indicating no evidence for practice level clustering. As these individual level factors are already included in the SEM, including clustering terms could lead to potential over adjustment.

### Recommendations for Future Research

Further studies are required to examine the applicability of the adapted model to other contexts. For example, research could investigate the applicability of the model to other providers of Patient Online (eg, Patient Access), specific platform (eg, mobiles, tablets, and computers), other settings (eg, hospitals), and other cities or countries.

Determining the factors that may influence the continuing use is important because long-term viability and eventual success of information technology depend on continued use [32,99-101]. Therefore, further primary studies and systematic reviews should be carried out to assess factors that affect the continuing use of ePHRs.

Further research is needed to explain the nonsignificant effect of SI demonstrated in this study. Previous studies demonstrated that the effect of SI depended on the type of processes of SI that people considered in their decisions (internalization, identification, and compliance) [37,84,102]. Thus, researchers may consider these 3 types of processes when assessing the effect of SI. Furthermore, researchers should develop new measures to assess the actual SI, such as the number of times a patient has been informed about the system by doctors, receptionists, friends, leaflets, posters, videos, and/or automated messages.

Although this study examined the effect of 4 moderators on most of the direct relationships, it did not examine their effects on the 2 indirect relationships (ie, EE→PE→BI and PPS→PE→BI). The effect of moderators on indirect relationships is called moderated mediation or conditional indirect effect [103-105]. To the best of our knowledge, the moderated mediating effect has not been examined in the context of ePHRs or CHITs. For this reason, future studies are required to test such an effect.

Finally, to increase the predictive power of the proposed model, future studies should consider adding other factors to the proposed model, such as patients' satisfaction, patient activation level, health status, perceived severity, perceived susceptibility, awareness of Patient Online, compatibility, and results demonstrability.

### Conclusions

This study examined the main factors that affected patients' use of ePHRs in England. The proposed model accounted for 48% of the variance in UB, indicating the existence of other, as yet unidentified, factors that influence adoption of ePHRs. Future studies should confirm the effect of the factors included in this model and identify additional factors. This study suggests that adoption rates are affected by key factors that should be taken into account for the successful implementation of ePHRs. For example, developers of ePHRs should involve patients in the process of designing the system to consider functions and features that fit patients' preferences and skills, thereby, create a useful and easy to use system.


### Acknowledgments

The authors would like to thank Dr Hamish Fraser for his help in the initial stages of the study, especially for his contribution to the development of the study design and setting up of links with practices.


### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Selection of theory.

[DOCX File, 69KB - jmir_v21i7e12373_app1.docx ]

### Multimedia Appendix 2

Conceptual definitions of constructs.

[DOCX File, 15KB - jmir_v21i7e12373_app2.docx ]

### Multimedia Appendix 3

Theoretical foundation.

[DOCX File, 79KB - jmir_v21i7e12373_app3.docx ]

### Multimedia Appendix 4

Characteristics of the GP practices.

[DOCX File, 16KB - jmir_v21i7e12373_app4.docx ]

### Multimedia Appendix 5

Measures of constructs.

[DOCX File, 18KB - jmir_v21i7e12373_app5.docx ]

### Multimedia Appendix 6

Questionnaire.

[DOCX File, 70KB - jmir_v21i7e12373_app6.docx ]

### Multimedia Appendix 7

Results of construct reliability.

[DOCX File, 12KB - jmir_v21i7e12373_app7.docx ]

### Multimedia Appendix 8

Results of convergent validity.

[DOCX File, 13KB - jmir_v21i7e12373_app8.docx ]

### Multimedia Appendix 9

Intercorrelation coefficients and squared roots of AVE.

[DOCX File, 14KB - jmir_v21i7e12373_app9.docx ]

### Multimedia Appendix 10

Item loadings and cross-loadings.

[DOCX File, 18KB - jmir_v21i7e12373_app10.docx ]

### References


1. Markle Foundation. 2003. The Personal Health Working Group: Final Report URL: https://www.markle.org/sites/default/files/final_phwg_report1.pdf [accessed 2019-06-03]
2. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to electronic health records: results from a survey of veterans. Med Care 2013 Mar;51(3 Suppl 1):S52-S56. [doi: 10.1097/MLR.0b013e31827808db] [Medline: 23407012]
3. Pai HH, Lau F, Barnett J, Jones S. Meeting the health information needs of prostate cancer patients using personal health records. Curr Oncol 2013 Dec;20(6):e561-e569 [FREE Full text] [doi: 10.3747/co.20.1584] [Medline: 24311957]
4. Tulu B, Trapp AC, Strong DM, Johnson SA, Hoque M, Trudel J, et al. An analysis of patient portal utilization: what can we learn about online patient behavior by examining portal click data? Health Syst 2017 Dec 19;5(1):66-79. [doi: 10.1057/hs.2015.5]
5. Laranjo L, Rodolfo I, Pereira AM, de Sá AB. Characteristics of innovators adopting a national personal health record in Portugal: cross-sectional study. JMIR Med Inform 2017 Oct 11;5(4):e37 [FREE Full text] [doi: 10.2196/medinform.7887] [Medline: 29021125]
6. Ose D, Kunz A, Pohlmann S, Hofmann H, Qreini M, Krisam J, et al. A personal electronic health record: study protocol of a feasibility study on implementation in a real-world health care setting. JMIR Res Protoc 2017 Mar 2;6(3):e33 [FREE Full text] [doi: 10.2196/resprot.6314] [Medline: 28254735]
7. Rice RC. CiteWeb. 2014. Healthcare Leaders' Lived Experiences Regarding the Implementation of Electronic Personal Health Records URL: http://citeweb.info/20141288190

8. Vermeir P, Degroote S, Vandijck D, van Tiggelen H, Peleman R, Verhaeghe R, et al. The patient perspective on the effects of medical record accessibility: a systematic review. Acta Clin Belg 2017 Jun;72(3):186-194. [doi: 10.1080/17843286.2016.1275375] [Medline: 28056665]
9. Alyami MA, Song YT. Removing Barriers in Using Personal Health Record Systems. In: Proceedings of the 15th International Conference on Computer and Information Science. 2016 Presented at: IEEE'16; June 26-29, 2016; Okayama, Japan. [doi: 10.1109/ICIS.2016.7550810]
10. Ochoa 3rd A, Kitayama K, Uijtdehaage S, Vermillion M, Eaton M, Carpio F, et al. Patient and provider perspectives on the potential value and use of a bilingual online patient portal in a Spanish-speaking safety-net population. J Am Med Inform Assoc 2017 Nov 1;24(6):1160-1164. [doi: 10.1093/jamia/ocx040] [Medline: 28460130]
11. Morton AA. Examining Acceptance of an Integrated Personal Health Record (PHR). Baltimore: University of Maryland; 2011.
12. Pagliari C, Detmer D, Singleton P. The Nuffield Trust. 2007. Electronic Personal Health Records: Emergence and Implications for the UK URL: https://www.nuffieldtrust.org.uk/files/2017-01/electronic-personal-health-records-uk-web-final.pdf [accessed 2019-06-10]
13. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. Br Med J 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]
14. National Health Service. 2016. Start Using GP Online Services URL: https://www.nhs.uk/using-the-nhs/nhs-services/gps/gp-online-services/ [accessed 2019-06-10]
15. National Health Service England. 2017. About GP Online Services URL: https://www.england.nhs.uk/gp-online-services/about-the-prog/ [accessed 2019-06-10]
16. Power BI. 2018. Patient Online Management Information (POMI): April 2017-May 2019 URL: https://tinyurl.com/ybptnhsu [accessed 2019-06-10]
17. Fung V, Ortiz E, Huang J, Fireman B, Miller R, Selby JV, et al. Early experiences with e-health services (1999-2002): promise, reality, and implications. Med Care 2006 May;44(5):491-496. [doi: 10.1097/01.mlr.0000207917.28988.dd] [Medline: 16641669]
18. Huygens MW, Vermeulen J, Friele RD, van Schayck OC, de Jong JD, de Witte LP. Internet services for communicating with the general practice: barely noticed and used by patients. Interact J Med Res 2015 Nov 24;4(4):e21 [FREE Full text] [doi: 10.2196/ijmr.4245] [Medline: 26601596]
19. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008;15(6):729-736 [FREE Full text] [doi: 10.1197/jamia.M2547] [Medline: 18756002]
20. Logue MD, Effken JA. Modeling factors that influence personal health records adoption. Comput Inform Nurs 2012 Jul;30(7):354-362. [doi: 10.1097/NXN.0b013e3182510717] [Medline: 22525046]
21. Najaftorkaman M, Ghapanchi AH, Talaei-Khoei A. Analysis of Research in Adoption of Person-Centred Healthcare Systems: The Case of Online Personal Health Record. In: Proceedings of the 25th Australasian Conference on Information Systems. 2014 Presented at: ACIS'14; December 8-10, 2014; Auckland, New Zealand p. 8-10 URL: http://openrepository.aut.ac.nz/bitstream/handle/10292/8104/acis20140_submission_55.pdf?sequence=1&isAllowed=y
22. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
23. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. Int J Med Inform 2019 Dec 21;126(2019):164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
24. Andrews L, Gajanayake R, Sahama T. The Australian general public's perceptions of having a personally controlled electronic health record (PCEHR). Int J Med Inform 2014 Dec;83(12):889-900. [doi: 10.1016/j.ijmedinf.2014.08.002] [Medline: 25200198]
25. Assadi V. MacSphere: McMaster University. 2013. Adoption of Integrated Personal Health Record Systems: A Self-determination Theory Perspective URL: https://macsphere.mcmaster.ca/bitstream/11375/13530/1/fulltext.pdf
26. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012 Nov 5;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]
27. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
28. Abramson EL, Patel V, Edwards A, Kaushal R. Consumer perspectives on personal health records: a 4-community study. Am J Manag Care 2014 Apr;20(4):287-296 [FREE Full text] [Medline: 24884860]
29. Nambisan P. Factors that impact patient web portal readiness (PWPR) among the underserved. Int J Med Inform 2017 Dec;102:62-70. [doi: 10.1016/j.ijmedinf.2017.03.004] [Medline: 28495349]
30. Ozok AA, Wu H, Gurses AP. Exploring patients' use intention of personal health record systems: implications for design. Int J Hum-Comput Int 2017 Jan 6;33(4):265-279. [doi: 10.1080/10447318.2016.1277637]

31. Razmak J, Bélanger C. Using the technology acceptance model to predict patient attitude toward personal health records in regional communities. Inform Tech People 2018 Apr 3;31(2):306-326. [doi: 10.1108/ITP-07-2016-0160]
32. Gebauer L, Söllner M, Leimeister JM. Towards Understanding the Formation of Continuous IT Use. In: Proceedings of the Thirty Fourth International Conference on Information Systems. 2013 Presented at: ICIS'13; December 15-18, 2013; Milano, Italy.
33. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. Infor Manag 2003 Jan;40(3):191-204. [doi: 10.1016/S0378-7206(01)00143-4]
34. Rodman DM. Meaningful Use: The Utilization of Patient Portals. Buffalo: D'Youville College; 2015.
35. Turner M, Kitchenham B, Brereton P, Charters S, Budgen D. Does the technology acceptance model predict actual use? A systematic literature review. Inf Softw Technol 2010 May;52(5):463-479 [FREE Full text] [doi: 10.1016/j.infsof.2009.11.005]
36. Bhattacherjee A. Social Science Research: Principles, Methods, and Practices. Scotts Valley, California , US: CreateSpace Independent Publishing Platform; 2012.
37. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
38. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
39. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
40. Baird A. Extending Adoption of Innovation Theory With Consumer Influence the Case of Personal Health Records (PHRS) and Patient Portals. Arizona: Arizona State University; 2012.
41. Dontje K, Corser WD, Holzman G. Understanding patient perceptions of the electronic personal health record. J Nurse Pract 2014 Nov;10(10):824-828. [doi: 10.1016/j.nurpra.2014.09.009]
42. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
43. Mishuris RG, Stewart M, Fix GM, Marcello T, McInnes DK, Hogan TP, et al. Barriers to patient portal access among veterans receiving home-based primary care: a qualitative study. Health Expect 2015 Dec;18(6):2296-2305 [FREE Full text] [doi: 10.1111/hex.12199] [Medline: 24816246]
44. Patel VN, Dhopeshwarkar RV, Edwards A, Barrón Y, Sparenborg J, Kaushal R. Consumer support for health information exchange and personal health records: a regional health information organization survey. J Med Syst 2012 Jun;36(3):1043-1052. [doi: 10.1007/s10916-010-9566-0] [Medline: 20703633]
45. Rao M. Factors affecting health care technology use in baby boomers: a quantitative study. California: Northcentral University; 2014:1.
46. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 3;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
47. Turner AN, Osterhage K, Joe J, Hartzler A, Lin L, Demiris G. Use of patient portals: personal health information management in older adults. Stud Health Technol Inform 2015;216:978. [doi: 10.1080/00221309.1994.9921212#] [Medline: 26262280]
48. Whetstone M, Goldsmith R. Factors influencing intention to use personal health records. Int J Pharm Healthc Mark 2009 Apr 3;3(1):8-25. [doi: 10.1108/17506120910948485]
49. Cho AH, Arar NH, Edelman DE, Hartwell PH, Oddone EZ, Yancy WS. Do diabetic veterans use the internet? Self-reported usage, skills, and interest in using My HealtheVet web portal. Telemed J E Health 2010 Jun;16(5):595-602. [doi: 10.1089/tmj.2009.0164] [Medline: 20575727]
50. Patel VN, Abramson E, Edwards AM, Cheung MA, Dhopeshwarkar RV, Kaushal R. Consumer attitudes toward personal health records in a beacon community. Am J Manag Care 2011 Apr;17(4):e104-e120 [FREE Full text] [Medline: 21774099]
51. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient race/ethnicity and shared medical record use among diabetes patients. Med Care 2012 May;50(5):434-440. [doi: 10.1097/MLR.0b013e318249d81b] [Medline: 22354209]
52. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]
53. Ronda MC, Dijkhorst-Oei L, Gorter KJ, Beulens JW, Rutten GE. Differences between diabetes patients who are interested or not in the use of a patient web portal. Diabetes Technol Ther 2013 Jul;15(7):556-563 [FREE Full text] [doi: 10.1089/dia.2013.0023] [Medline: 23777369]
54. Noblin AM, Wan TT, Fottler M. Intention to use a personal health record: a theoretical analysis using the technology acceptance model. Int J Health Technol Manag 2013;14(1/2):73-89. [doi: 10.1504/IJHTM.2013.055085]
55. van der Vaart R, Drossaert CH, Taal E, van de Laar MA. Patient preferences for a hospital-based rheumatology interactive health communication application and factors associated with these preferences. Rheumatology (Oxford) 2011 Sep;50(9):1618-1626. [doi: 10.1093/rheumatology/ker161] [Medline: 21551221]

56. Cahill JE, Lin L, LoBiondo-Wood G, Armstrong TS, Acquaye AA, Vera-Bolanos E, et al. Personal health records, symptoms, uncertainty, and mood in brain tumor patients. Neurooncol Pract 2014 Jun;1(2):64-70 [FREE Full text] [doi: 10.1093/nop/npu005] [Medline: 26034618]
57. Hibbard JH, Greene J. Who are we reaching through the patient portal: engaging the already engaged? Int J Pers Cent Med 2011;1(4):788-793. [doi: 10.5750/ijpcm.v1i4.152]
58. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in electronic health record patient portal use in nephrology clinics. Clin J Am Soc Nephrol 2015 Nov 6;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
59. Kim E, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009 Oct 27;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
60. Laugesen JD. MacSphere: McMaster University. 2013. Adoption of Electronic Personal Health Records by Chronic Disease Patients: Integrating Protection Motivation Theory and Task-Technology Fit URL: https://macsphere.mcmaster.ca/bitstream/11375/13343/1/fulltext.pdf
61. Davis SE, Osborn CY, Kripalani S, Goggins KM, Jackson GP. Health literacy, education levels, and patient portal usage during hospitalizations. AMIA Annu Symp Proc 2015;2015:1871-1880 [FREE Full text] [Medline: 26958286]
62. Martinez M, Baum A, Gomez SA, Gomez A, Luna D, González BF. Predictive variables of the use of personal health record: the hospital Italiano de Buenos Aires study. Stud Health Technol Inform 2013;192:1171. [doi: 10.3233/978-1-61499-289-9-1171] [Medline: 23920945]
63. Mikles SP, Mielenz TJ. Characteristics of electronic patient-provider messaging system utilisation in an urban health care organisation. J Innov Health Inform 2014 Dec 18;22(1):214-221 [FREE Full text] [doi: 10.14236/jhi.v22i1.75] [Medline: 25924551]
64. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis: A Global Perspective. Seventh Edition. Essex, UK: Pearson Education; 2010.
65. Kline RB. Principles and Practice of Structural Equation Modeling. Fourth Edition. New York, USA: Guilford Press; 2015.
66. Byrne BM. Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming. New York, USA: Routledge; 2016.
67. Kamarulzaman Y. ORCA - Cardiff University. 2006. The Adoption of Internet Shopping for Travel Services URL: https://orca.cf.ac.uk/55606/1/U584036.pdf
68. Chin WW. The partial least squares approach to structural equation modeling. In: Marcoulides GA, editor. Modern Methods for Business Research. Mahwah, USA: Lawrence Erlbaum Associates; 1998:295-336.
69. Field A. Discovering Statistics Using SPSS. Fifth Edition. London, UK: Sage Publication; 2017.
70. Schumacker RE, Lomax RG. A Beginner's Guide To Structural Equation Modeling. Third Edition. New York: Routledge; 2019.
71. Archer N, Cocosila M. Canadian patient perceptions of electronic personal health records: an empirical investigation. Commun Assoc Info Syst 2014;34(1):389-406. [doi: 10.17705/1CAIS.03420]
72. Assadi V, Hassanein K. Consumer adoption of personal health record systems: a self-determination theory perspective. J Med Internet Res 2017 Dec 27;19(7):e270 [FREE Full text] [doi: 10.2196/jmir.7721] [Medline: 28751301]
73. Lazard AJ, Watkins I, Mackert MS, Xie B, Stephens KK, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. J Am Med Inform Assoc 2016 Apr;23(e1):e157-e161 [FREE Full text] [doi: 10.1093/jamia/ocv174] [Medline: 26635314]
74. Liu CF, Tsai YC, Jang FL. Patients' acceptance towards a web-based personal health record system: an empirical study in Taiwan. Int J Environ Res Public Health 2013 Oct 17;10(10):5191-5208 [FREE Full text] [doi: 10.3390/ijerph10105191] [Medline: 24142185]
75. Majedi A. Semantic Scholar. 2014. Consumer Adoption of Personal Health Records URL: https://pdfs.semanticscholar.org/7d1e/27c3e792b2e5431a576e4293ca54f3d9a924.pdf
76. Alaiad A, Zhou L. The Moderating Effect of Demographic Characteristics on Home Healthcare Robots Adoption: An Empirical Study. In: Proceedings of the Tenth Midwest Association for Information Systems Conference. 2015 Presented at: MWAIS'15; May 14-15, 2015; Pittsburg, Kansas p. 14-15.
77. Rho MJ, Kim HS, Chung K, Choi IY. Factors influencing the acceptance of telemedicine for diabetes management. Clust Comput 2014 Mar 12;18(1):321-331. [doi: 10.1007/s10586-014-0356-1]
78. Noblin AM. Semantic Scholar. 2010. Intention To Use A Personal Health Record (PHR): A Cross Sectional View Of The Characteristics And Opinions Of Patients Of One Internal Medicine Practice URL: https://pdfs.semanticscholar.org/6fd2/9bd3c3b4cc3291297943977f887037bb0b1c.pdf
79. Wu H. Exploring Healthcare Consumer Acceptance of Personal Health Information Management Technology Through Personal Health Record Systems. Baltimore County: University of Maryland; 2013.
80. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 2;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]

81. Hsu C, Lee M, Su C. The role of privacy protection in healthcare information systems adoption. J Med Syst 2013 Oct;37(5):9966. [doi: 10.1007/s10916-013-9966-z] [Medline: 24014266]
82. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
83. Mathieson K. Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. Inf Syst Res 1991 Sep;2(3):173-191. [doi: 10.1287/isre.2.3.173]
84. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
85. Carlsson C, Carlsson J, Hyvonen K, Puhakainen J, Walden P. Adoption of Mobile Devices/Services - Searching for Answers with the UTAUT. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences. 2006 Presented at: HICSS'06; January 4-7, 2006; Kauia, HI, USA. [doi: 10.1109/HICSS.2006.38]
86. de Veer AJ, Peeters JM, Brabers AE, Schellevis FG, Rademakers JJ, Francke AL. Determinants of the intention to use e-health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
87. Daglish DE. MacSphere: McMaster University. 2013. Electronic Personal Health Records: A Matter of Trust URL: https://macsphere.mcmaster.ca/bitstream/11375/13325/1/fulltext.pdf
88. Arauwou JA. Older Adults' Perceptions of the UTAUT2 Factors Related to Intention to Use a Patient Portal for Engagement in Their Healthcare. California: Northcentral University; 2017.
89. Luque AE, van Keken A, Winters P, Keefer MC, Sanders M, Fiscella K. Barriers and facilitators of online patient portals to personal health records among persons living with HIV: formative research. JMIR Res Protoc 2013 Jan 22;2(1):e8 [FREE Full text] [doi: 10.2196/resprot.2302] [Medline: 23612564]
90. Hsieh H, Kuo Y, Wang S, Chuang B, Tsai C. A study of personal health record user's behavioral model based on the PMT and UTAUT integrative perspective. Int J Environ Res Public Health 2016 Dec 23;14(1):pii: E8 [FREE Full text] [doi: 10.3390/ijerph14010008] [Medline: 28025557]
91. Jian W, Syed-Abdul S, Sood SP, Lee P, Hsu M, Ho CH, et al. Factors influencing consumer adoption of USB-based personal health records in Taiwan. BMC Health Serv Res 2012 Aug 27;12:277 [FREE Full text] [doi: 10.1186/1472-6963-12-277] [Medline: 22925029]
92. Hoque R, Sorwar G. Understanding factors influencing the adoption of mhealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 Dec;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
93. Klein R. Internet-based patient-physician electronic communication applications: patient acceptance and trust. E-Service Journal 2007;5(2):27-52. [doi: 10.2979/esj.2007.5.2.27]
94. Bjerkan J, Hedlund M, Hellesø R. Patients' contribution to the development of a web-based plan for integrated care - a participatory design study. Inform Health Soc Care 2015 Mar;40(2):167-184. [doi: 10.3109/17538157.2014.907803] [Medline: 24786524]
95. Andrews JC, Shimp TA. Advertising, Promotion, and Other Aspects of Integrated Marketing Communications. Tenth Edition. Boston: Cengage Learning; 2017.
96. Kotler P, Armstrong G. Principles Of Marketing. Seventeenth Edition. Upper Saddle River, USA: Pearson; 2017.
97. Rogers EM. Diffusion of Innovations. Fifth Edition. New York, USA: Free Press; 2003.
98. Zikmund WG, Babin BJ, Carr JC, Griffin M. Business Research Methods. Ninth Edition. New York, USA: Cengage Learning; 2013.
99. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351-370. [doi: 10.2307/3250921]
100. Forquer HA, Christensen JL, Tan AS. Predicting continuance-findings from a longitudinal study of older adults using an ehealth newsletter. Health Commun 2014;29(9):937-946 [FREE Full text] [doi: 10.1080/10410236.2013.833580] [Medline: 24446900]
101. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011 Sep 30;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
102. Malhotra Y, Galletta DF. Extending the Technology Acceptance Model to Account for Social Influence: Theoretical Bases and Empirical Validation. In: Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999 Presented at: HICSS'99; January 5-8, 1999; Maui, HI, USA p. 5-8. [doi: 10.1109/HICSS.1999.772658]
103. Aguinis H, Edwards JR, Bradley KJ. Improving our understanding of moderation and mediation in strategic management research. Organ Res Methods 2016 Jan 27;20(4):665-685. [doi: 10.1177/1094428115627498]
104. Preacher KJ, Rucker DD, Hayes AF. Addressing moderated mediation hypotheses: theory, methods, and prescriptions. Multivariate Behav Res 2007;42(1):185-227. [doi: 10.1080/00273170701341316] [Medline: 26821081]
105. Sardeshmukh SR, Vandenberg RJ. Integrating moderation and mediation: a structural equation modeling approach. Organ Res Methods 2016 Jan 6;20(4):721-745. [doi: 10.1177/1094428115621609]

## Abbreviations

**AVE:** average variance extracted
**BI:** behavioral intention
**CHIT:** consumer health information technology
**EE:** effort expectancy
**ePHR:** electronic personal health record
**FC:** facilitating conditions
**GFI:** goodness-of-fit index
**GP:** general practice
**PE:** performance expectancy
**PPS:** perceived privacy and security
**RMSEA:** root mean square error of approximation
**SEM:** structural equation modeling
**SI:** social influence
**SRMR:** standardized root mean square residual
**UB:** use behavior
**UTAUT:** unified theory of acceptance and use of technology

*Edited by G Eysenbach; submitted 03.10.18; peer-reviewed by K Nazi, J Tavares, K Fiscella, B McMillan, M Nieroda; comments to author 02.02.19; revised version received 26.03.19; accepted 29.05.19; published 31.07.19.*

*Please cite as:*
*Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P*
*Factors Affecting Patients' Use of Electronic Personal Health Records in England: Cross-Sectional Study*
*J Med Internet Res 2019;21(7):e12373*
*URL: http://www.jmir.org/2019/7/e12373/*
*doi:10.2196/12373*
*PMID:31368442*

Review

# Telehealth Interventions for Improving Self-Management in Patients With Hemophilia: Scoping Review of Clinical Studies

Wenji Qian[1,2], BSc; Teddy Tai-Ning Lam[3], PharmD, PhD; Henry Hon Wai Lam[4], BSc, MSc; Chi-Kong Li[5,6], MBBS, MD, FRCPCH, FHKAM; Yin Ting Cheung[3], PhD

[1]School of Public Health, Fudan University, Shanghai, China

[2]Key Laboratory of Health Technology Assessment, National Health Commission of the People's Republic of China, Shanghai, China

[3]School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[4]The Hong Kong Haemophilia Society, Hong Kong, China (Hong Kong)

[5]Department of Paediatrics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[6]Paediatric Haematology & Oncology, Hong Kong Children's Hospital, Hong Kong, China (Hong Kong)

**Corresponding Author:**
Yin Ting Cheung, PhD
School of Pharmacy
Faculty of Medicine
The Chinese University of Hong Kong
8th Floor, Lo Kwee-Seong Integrated Biomedical Sciences Building
Hong Kong,
China (Hong Kong)
Phone: 852 3943 6833
Fax: 852 2603 5295
Email: yinting.cheung@cuhk.edu.hk

## Abstract

**Background:** The introduction of home therapy for hemophilia has empowered patients and their families to manage the disease more independently. However, self-management of hemophilia is demanding and complex. The uses of innovative interventions delivered by telehealth routes such as social media and Web-based and mobile apps, may help monitor bleeding events and promote the appropriate use of clotting factors among patients with hemophilia.

**Objective:** This scoping review aims to summarize the literature evaluating the effectiveness of telehealth interventions for improving health outcomes in patients with hemophilia and provides direction for future research.

**Methods:** A search was conducted in Ovid MEDLINE, EMBASE, and PubMed databases for studies that (1) focused on patients with hemophilia A or B; (2) tested the use of remote telehealth interventions via the internet, wireless, satellite, telephone, and mobile phone media on patients and caregivers; and (3) reported on at least one of the following patient-/caregiver-focused outcomes related to empowering patients/caregivers to be active decision makers in the emotional, social, and medical management of the illness: quality of life, monitoring of bleeding episodes, joint damage or other measures of functional status, medication adherence, and patients' knowledge. Implementation outcomes (user metrics, cost saving, and accuracy of electronic records) were also evaluated. Reviews, commentaries, and case reports comprising ≤10 cases were excluded.

**Results:** Sixteen articles fulfilled the inclusion criteria. The majority of the interventions (10/16, 62%) evaluated both implementation outcomes and patient-/caregiver-focused outcomes. User performance and accuracy and comprehensiveness of electronic records were also measured in most studies (4/16, 87%). The components of the interventions were rather homogenous and typically involved electronic logging and reminders for prophylactic infusions, reporting of spontaneous and traumatic bleeding events, monitoring of infusion product usage and home inventory, and real-time communication with health care professionals and hemophilia clinics. Telemedicine-supported education and information interventions seemed to be particularly effective among adolescent and young adult patients. Although the patients reported improvements in their health-related quality of life and perception of illness, telemonitoring devices did not appear to have a significant effect on quantifiable health outcomes such as joint health. Longitudinal studies seemed to suggest that the response and adherence rates to recording decreased over time.

**Conclusions:** Preliminary evidence from this review suggests that telehealth-delivered interventions could feasibly improve patients' adherence to medication use and promote independence in disease management. Given the complexity and resources involved in developing a mature and established system, support from a dedicated network of hemophilia specialists and data managers will be required to maintain the technology, improve adherence to prophylactic treatment and recording, and validate the electronic data locally.

***(J Med Internet Res 2019;21(7):e12340)*** doi:10.2196/12340

**KEYWORDS**

hemophilia; clotting factors; adherence; self-management; telehealth

## Introduction

Hemophilia refers to a rare group of X-linked recessive hemorrhagic diseases that often require complex disease care and treatment. Appropriate management is needed to reduce the bleeding episodes and long-term complications of bleeding, such as chronic arthropathy and intracerebral hemorrhage, as well as the frequency of hospitalization and absenteeism from school or work [1]. Fortunately, the introduction of home therapy for hemophilia has empowered patients and their families to manage the disease more independently. However, the self-management of hemophilia is demanding, including the recognition of bleeding, adherence to self-administration of prophylactic infusions, recording of bleeding events, and management of a home inventory of medications.

Strategies are needed to improve the health outcomes and self-efficacy of patients with hemophilia. To address these needs, platforms involving different types of technology have been implemented to promote health education and good protective health behavior. The Health Resources and Services Administration defines the term "telehealth" as the use of electronic information and telecommunications technologies to support and promote long-distance clinical health care, health-related education for patients and professionals, public health, and health administration [2]. For the purpose of this review, we have broadly defined "telemedicine" or "telehealth intervention" as interventions that use telecommunications technology to facilitate the remote delivery of health care services and clinical information [2,3]. These interventions can be synchronous or asynchronous and include any information and technology-based strategies for connecting health care professionals and patients through video conferencing, e-mail, remote electronic monitoring equipment, social network apps, and internet portals [3]. By nature, telehealth interventions include interactive telemedicine services that facilitate concurrent interactions among patients, caregivers, and clinicians, and the remote monitoring of patients' health statuses using telehealth equipment and "store-and-forward telemedicine," which involves the transmission of disease-related data such as medical images, bleeding, episodes, and biological measures [4]. As adherence to the recording of bleeding episodes and infusion plays an important role in treatment outcomes, applications of telehealth intervention in hemophilia include the use of an electronic device to collect information on bleeding episodes and real-time transmission to the hemophilia clinic or the use of videoconferencing to educate patients on infusion techniques in a remote setting [4].

The literature includes a preponderance of reviews that discuss the benefits and challenges of interventions in the management of chronic diseases delivered via telehealth [4-6]. Emerging evidence supports the use of technology to facilitate the self-management of complex and demanding treatments for chronic diseases. The uses of innovative interventions delivered by telehealth routes, such as social media, mobile apps, and teleconferences, may help monitor bleeding events and promote adherence in prophylactic infusion of clotting factors among patients with hemophilia through built-in alarms and reminders [4]. This delivery route may also be harnessed to increase patients' motivation for participating in self-care activities and protective health behaviors in terms of the recommended diet and exercise. Although previous systematic reviews have focused on the applicability of interventions delivered by telehealth for patients with chronic conditions, to our knowledge, none have involved the population of patients with hemophilia. This review aims to summarize the literature evaluating the effectiveness of telehealth interventions for improving health outcomes in patients with hemophilia and to provide directions for future research.

## Methods

### Overall Approach

In the initial phase of this study, the original protocol was developed as a systematic review to quantitatively address the feasibility, appropriateness, and effectiveness of telehealth intervention in patients with hemophilia. However, a preliminary search revealed high heterogeneity in the assessment of health outcomes across studies with limited sample sizes and generalizability. Consensus among the investigators resulted in a protocol modification to perform a scoping review instead. We rationalized that this is a reasonable decision because scoping reviews follow a narrative synthesis approach, which is more appropriate for summarizing review studies that vary in methodological approaches.

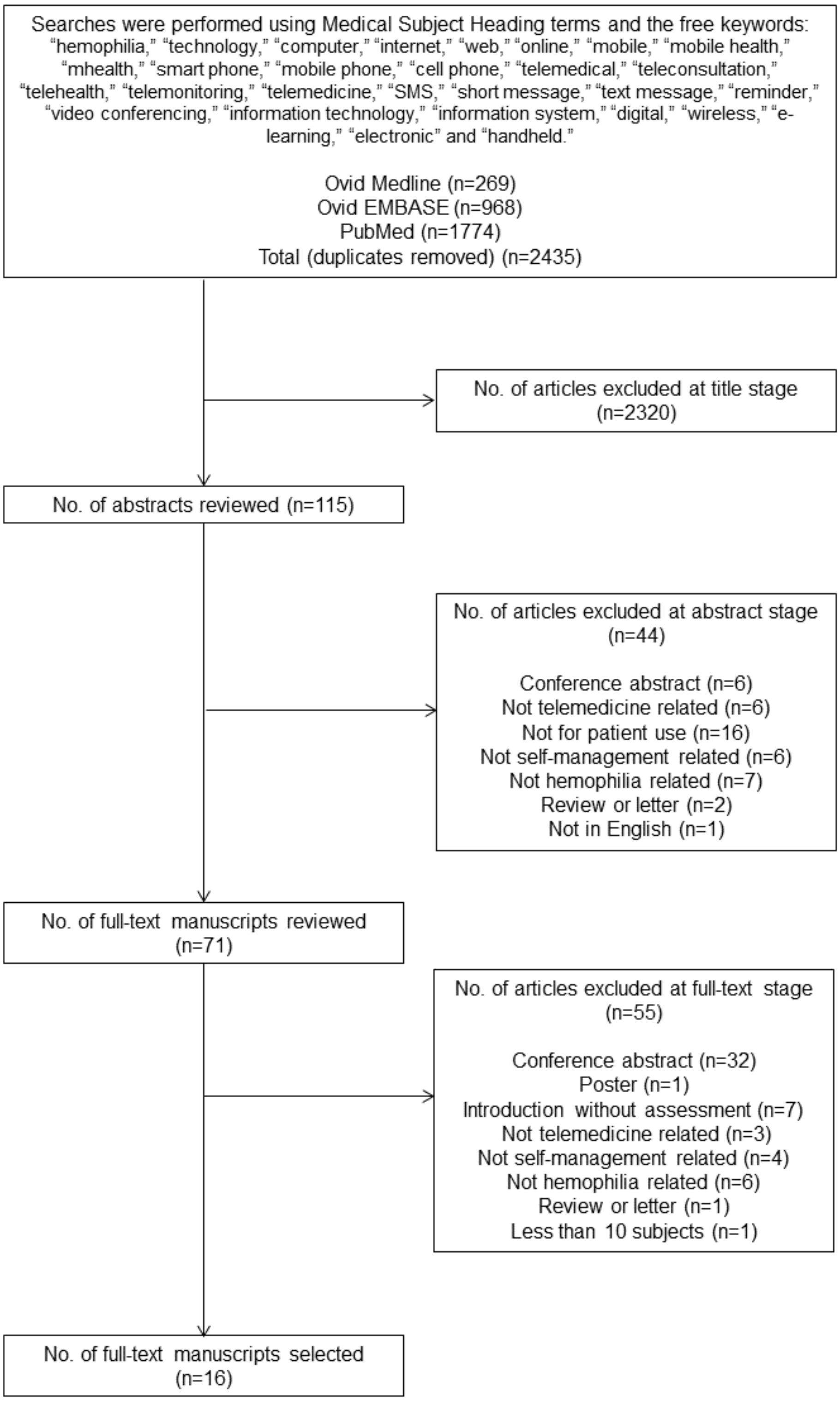
We followed the steps recommended by Arskey and O'Malley [7] to perform the review: (1) identify the research questions; (2) identify relevant studies; (3) select studies; (4) chart the data; and (5) collate, summarize, and report the results of the included studies. To ensure transparency and complete reporting, we followed the PRISMA extension (PRISMA-ScR) for scoping reviews [8].

### Search Procedure

The Ovid MEDLINE, EMBASE and PubMed databases were searched in July 2018 using the following combination of terms:

"hemophilia," "technology," "computer," "internet," "web," "online," "mobile," "mobile health," "mhealth," "smart phone," "mobile phone," "cell phone," "telemedical," "teleconsultation," "telehealth," "telemonitoring," "telemedicine," "SMS," "short message," "text message," "reminder," "video conferencing," "information technology," "information system," "digital," "wireless," "e-learning," "electronic" and "handheld." Only English-language articles published before July 31, 2018, were reviewed. The search strategies comprised Medical Subject Headings, keywords, and text words related to bleeding disorders and telehealth. The complete search strategies are provided in Figure 1. We also manually searched the reference lists of the retrieved manuscripts.

**Figure 1.** Flowchart of the literature search.




The inclusion and exclusion criteria were as follows: studies focusing on patients with hemophilia A or B; studies testing the use of remote telehealth interventions via the internet, wireless, satellite, telephone, and mobile phone media on the patient or caregiver of the patient; and studies reporting on at least one of the following primary outcomes related to empowering patients or caregivers to be active decision makers in the emotional, social, or medical management of their illness or the child's illness. These outcomes were identified from the literature as critical for improving patient efficacy in the management of hemophilia [9] as well as by clinical consensus from the team of investigators comprising methodologists (WQ and YC), a hematologist (CL), a pharmacist (TL), and a representative of a support group for patients with hemophilia (HL). Examples of such outcomes included quality of life; monitoring and adherence to the recording of bleeding episodes; adherence to prophylactic infusion; joint damage or other measures of functional status; medication adherence; patient knowledge; and patients' perceived value, acceptance, and satisfaction with the intervention. Secondary outcomes related to the development and implementation of the intervention were also evaluated. These outcomes included, but were not limited to, user metrics, the comparison of reliability and accuracy between electronic and paper records, and impact of intervention on cost and resource saving.

The exclusion criteria were as follows: (1) reviews, commentaries, conference proceedings, dissertation reports, or case reports comprising 10 or fewer cases and (2) studies that did not describe the basic quantitative or qualitative research methodology such as data collection methods, clinical assessment methods and definitions, and analytic or reporting strategies.

### Data Extraction and Analysis

The search results were reviewed on three sequential levels: (1) In the initial "title stage," the article titles were screened to exclude studies that clearly did not fulfil the inclusion criteria outlined in this review. (2) In the "abstract stage," the abstracts of articles that passed the "title stage" were reviewed. (3) In the final "full-text stage," the remaining articles were examined to ensure that they fulfilled the inclusion/exclusion criteria. The screening and data extraction phases were conducted by the investigators (WQ and YC) independently. The list of included studies and summary of data prepared by the two investigators were then compared, and disagreements were resolved through discussion and involvement of a third investigator (TL). The study characteristics were systematically abstracted using a structured data collection form that included the following parameters: the country in which the study was conducted, year of publication, study design, sample size, patient characteristics, description of intervention, assessment outcomes, and tools and conclusion.

### Quality Assessment

Although quality assessment of the included studies is generally not required for a scoping review, assessment of the methodological limitations was evaluated to establish the quality of existing evidence and address variation in the study approaches. Two reviewers assessed the methodological quality of the observational studies and controlled trials (WQ and YC) using the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies (EPHPP) [10]. Any discrepancies in the ratings were resolved by discussion. Interrater reliability was calculated using the Cohen kappa statistic to ascertain the agreement between the scores of the two reviewers in terms of each criterion of the EPHPP [11]. The EPHPP determines the quality of each study based on six criteria: selection bias, study design, confounders, blinding, data collection methods, and withdrawals/dropouts. Each individual component was ranked as strong, moderate, or weak according to the EPHPP. Finally, the sum of all six criteria was calculated to construct a global rating that represents the overall robustness and quality of the study methodology. Two additional EPHPP components—integrity and consistency of the intervention and the appropriateness of analytical methods—are independent scales that are not included in the overall grading [10].

## Results

### General Characteristics

The results of the literature search are depicted in Figure 1. The search resulted in 2435 titles. After comparing the titles and abstracts against the eligibility criteria, 2364 articles were excluded, leaving 71 articles for full-text screening. After an independent review of the full texts, 16 articles fulfilled the inclusion criteria [12-27]. The general characteristics of these studies are presented in Table 1.

Of the 16 articles, three (19%) described randomized controlled trials (RCTs) [14,25,26], two (12%) described qualitative studies [15,27], and the remaining (n=11, 69%) described observational studies [12,13,16-24]. Eight studies (50%) were conducted in North America [14,15,19,20,22,23,26,27]; five (31%), in Europe [16-18,24,25]; two (12%), in the United Kingdom [12,13]; and one (6%), in Australia [21].

The sample sizes of the studies were generally small, ranging from 10 to 50 patients. Broderick et al [21] investigated the response rate and expenses incurred from implementing a short message service (SMS) intervention in a cohort of 104 patients with hemophilia in Australia [21]. One large-scale study recruited 2683 patients from different regions of the United Kingdom to evaluate usage metrics and compliance with a nationwide electronic recording platform [13]. The majority of studies involved patients with severe hemophilia who required prophylactic infusion therapy. The age distributions varied widely across the studies, which include both adult patients and caregivers of pediatric or adolescent patients. Two studies specifically targeted the delivery of educational resources regarding self-management of adolescents between the ages of 12 and 18 years [26,27].

**Table 1.** Characteristics of patients in the included studies.

| Study and year | Country | Design[a] | Sample size (rate) | Disease type (%) | Disease severity | Treatment[b] (%) | Age (years), proportion or as indicated | Education level |
|---|---|---|---|---|---|---|---|---|
| Collins et al, 2003 [12] | United Kingdom | OS[c] longitudinal | Total: 10 | Hem[d] | NR[e] | NR | 12-17 years: 70%, 30-45 years: 20%, >50 years: 10% | NR |
| Walker et al, 2004[f] [14] | Canada | RCT[g] | Total: 41, test: 22, control: 19, rate: 60% | HemA[h]: 93%, HemB[i]: 7% | Severe: 100% | Prophylaxis: 59% | Median: 25 years (IQR[j]: 15-42 years); <18 years: 63%, ≥18: 37% | NR |
| Arnold et al, 2005 [15] | Canada | OS/ qualitative | Total: 20 | HemA, HemB | Severe: 100% | NR | Range: 6-43 | NR |
| Pattacini et al, 2009 [16] | Italy | OS | Total: 50 | NR | NR | NR | NR | NR |
| Petrini et al, 2009 [17] | Denmark, Finland, Norway, Sweden | OS | Total: 57 | HemA: 100% | Moderate: 5%, severe: 95% | All on ReFacto, on demand: 19%, prophylaxis: 81% | Median (range): 17.5 (0.4-57.2) | NR |
| Mondorf et al, 2009 [18] | Germany | OS | Total: 10 | HemA | Severe: 100% | On demand: 10%, prophylaxis: 90% | Median (range): 31 (11-48) | NR |
| Vallee-Smejda et al, 2009 [19] | Canada | OS | Total: 18, rate: 47% | Hem | Mild, moderate, severe | NR | Range: 2-67; 16-20: 6%, 21-40: 44%, 41-60: 50% | High school: 33%, college: 22%, undergraduate: 28%, doctorate: 6% |
| Leone et al, 2011 [20] | United States | OS | Total: 52, rate: 100% | HemA, HemB | NR | Prophylaxis: 100% | All >5 years | NR |
| Broderick et al, 2012 [21] | Australia | OS | Total: 104 | HemA, HemB | Moderate, severe | NR | Mean (SD): 9.5 (4.0); range: 4-18 | NR |
| Mulders et al, 2012 [25] | The Netherlands | RCT | Total: 30, test: 16, control: 14, rate: 76% | HemA: 77%, HemB: 23% | Mild: 3%, moderate: 36%, severe: 60% | Prophylaxis: 40%, on demand: 26%, both: 34% | Median (range): 33.5 (17-67) | Elementary: 6%, secondary: 43%, vocational advanced: 36%, academic: 4 (13%) |
| Young et al, 2012 [22] | United States | OS | Total: 52 | HemA, HemB | ≥4 bleeding events in the prior 3 months | On demand: 71%, ITT[k]: 10%, prophylaxis: 23% | <18: 48%, ≥18: 52% | NR |
| Breakey et al, 2013 [27] | Canada | OS/qualitative | Total: 18, rate: 80% | HemA: 67%, HemB: 22%, Unsure: 11% | Mild: 17%, moderate:17%, severe: 66% | Prophylaxis: 78% | Median (range): 15.5 (13-18) | English speaking in grade 9-12: 67%, French speaking in grade 3-5: 33% |
| Sholapur et al, 2013 [23] | Canada | OS | Total: 51, test: 23, others: 28, rate: 62% | NR | NR | NR | <16: 39%, ≥16: 61% | NR |
| Breakey et al, 2014 [26] | Canada | RCT | Total: 29, test: 16, control: 13 | HemA: 62%, HemB: 31%, Unsure: 7% | Mild: 21%, moderate: 24%, severe: 55% | Prophylaxis: 76%, inhibitor: 4% | Mean (SD): 15.9 (1.34); range: 13-18 | Grade 10 (median) |
| Hay et al, 2017 [13] | United Kingdom | OS | Total: 2683 | HemA: 78%, HemB: 15%, Others: 7% | Mild: 9%, moderate 10%, severe: 81% | NR | NR | NR |
| Cuesta-Barriuso et al, 2018 [24] | Spain | OS longitudinal | Total: 43 | HemA: 86%, HemB: 14% | Severe: 91% | Prophylaxis: 100% | Mean (SD): 25.8 (10.3) | Primary: 7%, secondary: 44%, university: 49% |

[a]Design: Due to the heterogeneity of methodologies across studies, study designs are classified simply as either randomized controlled trial (RCT) or observational studies for non-RCT studies. Studies that utilized qualitative methods (eg, structured interviews) are specified.

[b]Treatment percentages may not add up to 100% because respondents might have indicated the use of multiple agents.

[c]OS: observational study.

[d]Hem: hemophilia.

[e]NR: not reported.

[f]Patient sample in Arnold, 2005 [15] is a subset of patients from Walker, 2004 [14].

[g]RCT: randomized controlled trial.

[h]HemA: hemophilia A.

[i]HemB: hemophilia B.

[j]IQR: interquartile range.

[k]ITT: immune tolerance treatment.

## Quality of Studies

The assessment of study quality received an interrater agreement *k* of 0.78. Most studies received a "weak" (7/16, 44%) or "moderate" (6/16, 38%) rating, although three studies were considered to be of "strong" methodological quality (Table 2). When considering the individual components of quality, most studies were considered "weak" if they were cross-sectional and single arm in nature and "moderate" or "strong" if they had a controlled trial or cohort design or a pre- and postintervention assessment. Adjustment for confounders is generally not applicable to studies involving only a single arm or those with poorly characterized clinical descriptions of the cohort. A minority of studies (n=2) reported assessments of outcomes blinded to the appropriate members of the research team. Half of the studies applied objective measures (eg, usage metrics, number of diary entries, and cost data) or cited data on the psychometric properties or development methodology behind their instruments of outcome assessments, while others were rated as "weak" if a satisfaction survey was the only mode of evaluation.

**Table 2.** Quality assessment of the included studies. Methodological quality scores of included studies are scored using the "Quality Assessment Tool for Quantitative Studies" developed by the Effective Public Health Practice Project [10].

| Study and year | Selection bias | Study design | Confounders | Blinding | Data collection methods | Withdrawals and dropouts | Intervention integrity | Analyses | Global |
|---|---|---|---|---|---|---|---|---|---|
| Collins et al, 2003 [12] | Weak | Moderate | N/A[a] | Weak | Weak | Moderate | Moderate | Weak | Weak |
| Walker et al, 2004 [14] | Moderate | Strong | Strong | Moderate | Strong | Strong | Moderate | Strong | Strong |
| Arnold et al, 2005 [15] | Weak | Weak | Weak | Weak | Moderate | N/A | Moderate | Moderate | Weak |
| Pattacini et al, 2009 [16] | Weak | Weak | N/A | Weak | Weak | N/A | Moderate | Moderate | Weak |
| Petrini et al, 2009[b] [17] | Moderate | Weak | N/A | Weak | Weak | Moderate | Moderate | Weak | Weak |
| Mondorf et al, 2009 [18] | Weak | Weak | N/A | Weak | Weak | Moderate | Weak | Weak | Weak |
| Vallee-Smejda et al, 2009 [19] | Weak | Moderate | N/A | Weak | Weak | Weak | Moderate | Weak | Weak |
| Leone et al, 2011 [20] | Moderate | Moderate | N/A | Weak | Weak | Strong | Strong | Weak | Moderate |
| Broderick et al, 2012 [21] | Moderate | Moderate | Moderate | Weak | Strong | Strong | Moderate | Weak | Moderate |
| Mulders et al, 2012 [25] | Moderate | Strong | Moderate | Moderate | Moderate | Strong | Moderate | Moderate | Moderate |
| Young et al, 2012 [22] | Weak | Moderate | N/A | Weak | Moderate | Moderate | Moderate | Moderate | Moderate |
| Breakey et al, 2013 [27] | Moderate | Moderate | N/A | Weak | Moderate | N/A | Moderate | Moderate | Moderate |
| Sholapur et al, 2013 [23] | Weak | Weak | Weak | Weak | Weak | N/A | Weak | Moderate | Weak |
| Breakey et al, 2014 [26] | Moderate | Strong | Moderate | Weak | Strong | Strong | Moderate | Strong | Strong |
| Hay et al, 2017 [13] | Moderate | Weak | Strong | Weak | Strong | N/A | Moderate | Strong | Strong |
| Cuesta-Barriuso et al, 2018 [24] | Moderate | Moderate | N/A | Weak | Strong | Moderate | Strong | Strong | Moderate |

[a]N/A: not applicable.

## Intervention Characteristics and Outcomes

Intervention characteristics and outcomes of the included studies are presented in Table 3. The majority of the interventions (10/16, 62%) evaluated both implementation outcomes and patient-/caregiver-focused outcomes. The most commonly adopted patient-/caregiver-focused outcomes included joint health, adherence to prophylactic treatment, health-related quality of life, and self-management skills. User performance and accuracy and comprehensiveness of electronic records were also measured in most studies (14/16, 87%).

The components of the interventions were rather homogenous and typically involved electronic logging and reminders for prophylactic infusions, electronic reporting of spontaneous and traumatic bleeding events, electronic monitoring of infusion product usage and home inventory, and electronic real-time communication with health care professionals and hemophilia clinics. The use of electronic diaries in the form of handheld computers and Web-based apps was the most common mode of intervention. However, the findings were rather inconsistent and depended on the outcome of interest. Narrative syntheses found that electronic diaries seemed to improve patient adherence and accuracy when recording bleeding episodes and infusion logs [14,15,19,24]. Using a validated scale for measuring patient's adherence to prophylactic treatment, Cuesta-Barriuso et al [24] reported a statistically significant decrease in adherence problems in patients at 12 months from the initiation of the Medtep Hemophilia online platform (baseline mean score: 44.6 [SD 10.4]; 12-month postintervention mean score: 33.6 [SD 6.5]; $P<.001$). The same study found a statistically significant improvement in most quality of life components on the Short Form-36, such as general health (baseline mean score: 48.4 [SD 9.3]; 12-month score: 57.1 [SD 5.6]; $P<.001$) and body pain (baseline mean score: 49.9 [SD 7.6]; 12-month score 52.8 [SD 6.3]; $P<.05$) [24]. Vallee-Smejda et al [19] demonstrated that the amount of completed additional fields nearly doubled with the use of an electronic diary, indicating improvement in the completeness of data entry. In terms of accuracy, one study reported a reasonably good agreement of 93% between the electronic records and paper diaries [22]. Leone et al [20] also emphasized how patients viewed the importance of the image-taking features of their monitoring devices for the purpose of capturing photographs of their bleeding joints.

However, longitudinal studies seemed to suggest that the rate of adherence to electronic reporting decreased over time [13,17]. For example, the response rate of reporting bleeding events through an SMS intervention decreased from approximately 95% at initiation to 85% after 1 year [21]. Petrini et al [17] reported that a decrease in usage of electronic recording was largely attributed to technical problems and the challenges involved in correcting errors that required contact with the primary clinics. Although the patients reported improvements in their health-related quality of life and perception of illness, telemonitoring devices did not appear to have a significant effect on quantifiable health outcomes such as joint health [24]. The assessment of such long-term indicators may have been limited by the short time horizons of the included studies, which ranged from only 8 weeks to 1 year. Patients who relied solely on on-demand treatments did not report any benefit of an electronic documentation system [18].

Three studies that focused on the provision of disease-related information and practical skills regarding the management of hemophilia yielded promising outcomes [25-27]. Specifically, these studies implemented a robust methodology with appropriate randomization strategy and analytic methods that account for confounding factors. Mulders et al [25] reported that patients who engaged in a 1-month educational electronic learning program demonstrated significantly higher levels of theoretical knowledge on hemophilia, bleeding treatment, and complications of treatment than control subjects (mean score: 75% vs 54%; $P<.001$) as well as better skills in the intravenous injection of clotting factor concentrates ($P<.001$). Telemedicine-supported education and information interventions seemed to be particularly effective among adolescent patients [26,27]. Breakey et al [26] evaluated the feasibility of an internet-based self-management and transitional care program for adolescents with hemophilia. They found that, compared to controls, adolescents in the intervention arm showed a significant improvement in disease-specific knowledge ($P=.004$), self-efficacy ($P=.007$), and transition preparedness ($P=.046$) using a structured questionnaire [26]. In addition, adherence to the completion of the final online outcome measures revealed that 17 of 18 (94%) teenagers successfully completed all the postprogram outcome measures.

**Table 3.** Characteristics of telehealth interventions and main outcomes.

| Study | Intervention features | Patient-/caregiver-focused outcomes | Implementation and intervention-focused outcomes | Findings |
|---|---|---|---|---|
| Collins et al, 2003 [12] | • Advoy.com (internet-based electronic patient treatment record and communication system)<br>• Documented bleeding event, infusion log, and symptoms<br>• Triggers and alerts can be set by clinicians<br>• Patient contacted by phone<br>• Duration: 8 weeks | • Perceived value of intervention | • Comparison of electronic and previous paper treatment records | • Reported electronic recording to be easier for treatment log: 8 patients (80%)<br>• Reported electronic recording to be worse for treatment log: 2 patients (20%)<br>• Reported electronic treatment log to be more accurate: 9 patients (90%) |
| Walker et al, 2004 [14] | • Dialog: Electronic diary<br>• Handheld computer<br>• Documented infusion, bleeding event, symptoms, and productivity<br>• Data transmitted to clinic<br>• Bar code reader for medications<br>• Single reminder phone call at the end of every month<br>• Preintervention training for patients<br>• Duration: 6 months | —[a] | • Usage metrics<br>• Accuracy and comprehensiveness of data | • 86.2% of infusions by patients in the intervention group were adherent with the data submission schedule, as compared to only 48.3% in the control group<br>• The time intervals between infusions and the receipt of data were shorter in the intervention group (median 0.25 days) as compared to the control group (25 days).<br>• Accuracy of data was similar for both methods.<br>• Reminder phone calls by the clinic made less frequently to the intervention group (median: 1 time/month) as compared to the control group (median: 5 times/month) |
| Arnold et al, 2005 [15] | • Dialog: Electronic diary<br>• Handheld computer (as in [14]) | • Patients' preferred choice of recording method: paper diaries versus handheld computers through semistructured interviews | • Usage metrics | • All patients preferred using handheld computers to using paper diaries<br>• 90% believed that their adherence to record keeping had improved using handheld computers |
| Pattacini et al, 2009 [16] | • "xl_Emofilia" (Web-based app)<br>• Record bleeding events and home infusions<br>• Collaborating sites have access to patient data<br>• Preintervention training for patients | • Level of patient satisfaction | • Usage metrics<br>• Degree of accuracy from validation | • 825 log-ins made<br>• 105 bleeding episodes or traumatic events recorded<br>• Degree of accuracy: 80% in the first month and 95% in the subsequent period<br>• High degree of acceptance among patients |
| Petrini et al, 2009[b] [17] | • Electronic diary<br>• Handheld computer<br>• Documented bleeding event | — | • Usage metrics | • Adherence was lower than expected, with ≤50% reporting accurately during the entire study period.<br>• Some patients reported a large number of infusions from a long time period, on one day.<br>• Technical problems were a major contributing factor to poor adherence.<br>• Examination of the diary data revealed useful information on the management of bleeding episodes. |

| Study | Intervention features | Patient-/caregiver-focused outcomes | Implementation and intervention-focused outcomes | Findings |
|---|---|---|---|---|
| Mondorf et al, 2009 [18] | • Haemoassist (handheld electronic diary)<br>• Documented bleeding event and factor infusion<br>• Access to patient data by clinicians<br>• Alert function to warn patients and physicians of critical clinical events<br>• Duration: 3-12 months | • Patient satisfaction | • Feasibility<br>• Usage metrics | • Very satisfied: 2 patients<br>• Moderately satisfied: 4 patients<br>• Not satisfied: 1 patient<br>• Nine patients continued the electronic documentation after the study.<br>• On-demand patients do not see any benefit in an electronic documentation system. |
| Vallee-Smejda et al, 2009 [19] | • Advoy.com (internet-based electronic patient treatment record and communication system)<br>• Duration: 6 months | • Adherence to recording<br>• Patient satisfaction | • Usage metrics | • Significantly more patients (29.4%) indicated satisfaction with electronic recording, as compared with paper records (6.7%).<br>• Electronic recording significantly improved patient adherence in recording mandatory treatment information.<br>• Electronic recording resulted in providing additional health data. |
| Leone et al, 2011 [20] | • HeliTrax System handheld monitoring device<br>• Duration: 3 months | • Perceived value of intervention | • Usage metrics<br>• Ease of use<br>• Proficiency of generated reports | • 86% of patients rated higher value for electronic recording over traditional paper logging.<br>• Approximately 90% of patients rated the ease of tracking as good or excellent.<br>• Approximately 80% of patients rated the picture-taking capability and importance of that feature as good. |
| Broderick et al, 2012 [21] | • Weekly SMS[c] to monitor incidence of bleeding episodes<br>• Duration: 52 weeks | — | • Response rate (proportion of weeks in which participants responded to the SMS)<br>• Cost | • Response rate: 86.8%<br>• Small but significant decrease in response rate over 52 weeks<br>• Use of SMS is associated with high response rates and minimal expense and intrusion. |
| Mulders et al, 2012 [25] | • E-learning (online course)<br>• Interactive multimedia program<br>• Education on home treatment of hemophilia<br>• Duration: 1 month | • Knowledge on home treatment<br>• Practical skills: self-injection<br>• Self-efficacy | | • Significantly higher levels of theoretical knowledge and practical skills in the intervention group, as compared to the control group. No group difference in self-efficacy |
| Young et al, 2012 [22] | • Electronic diary<br>• Internet-based entries submitted in real time<br>• Documented bleeding event, productivity, HRQoL[d], and pain assessment<br>• Weekly contact by patient support liaison personnel | — | • Usage metrics<br>• Degree of accuracy from validation | • Adults: 1364 paper and electronic diary days<br>• Caregivers: 1165 paper and electronic diary days<br>• Exact agreement observed between electronic and paper records for 93% of the HRQoL scores reviewed |

| Study | Intervention features | Patient-/caregiver-focused outcomes | Implementation and intervention-focused outcomes | Findings |
|---|---|---|---|---|
| Breakey et al, 2013 [27] | • "Teens Taking Charge: Managing Hemophilia Online" (interactive website to help patients transit from pediatric to adult care)<br>• Include hemophilia-specific education, self-managementstrategies, images, interactive animations, quizzes, and a glossary | • Participants' satisfaction<br>• Qualitative usability testing approach with audio-taped observation and semistructured interviews | • User performance | • Adolescent patients responded positively to the content and appearance of the website.<br>• Adolescent patients felt that it was easy to navigate and understand<br>• The multimedia components (videos, animations, and quizzes) enriched their experience. |
| Sholapur et al, 2013 [23] | • EZ-Log Web Client (electronic diary) | • Identify strengths and challenges of traditional vs electronic diaries | • Usage metrics | • Advantages: ease of use, improved accuracy, and less time consuming<br>• Disadvantage: Technical errors and inability to make corrections that require contact with the clinic<br>• Suggestions: Saving infusion history, incorporating barcode scanners |
| Breakey et al, 2014 [26] | • Eight-module online program<br>• Interactive content<br>• Animations, illustrations, and knowledge quizzes<br>• Duration: 8-10 weeks | • Disease-specific knowledge<br>• HRQoL<br>• Self-efficacy<br>• Self-management ability and transition readiness<br>• Program satisfaction | • Not applicable | • Significant improvement in disease-specific knowledge, self-efficacy, and transition preparedness in the intervention group<br>• Program informative, comprehensive, and easy to use |
| Hay et al, 2017 [13] | • Haemtrack (electronic home treatment diary) interfaces with the UK Haemophilia Centre Information System and theNational Haemophilia Database<br>• Documented bleeding event, infusion log, pain assessment, and outcome | • Adherence to electronic recording | • Usage metrics | • Electronic diary used by 90% of participating hemophilia treatment centers<br>• 72% (1923/2683) of patients used electronic diary, entering >17,000 treatments per month<br>• Adherence to reporting varied, and 55% of patients reported ≥75% of expected factor usage.<br>• No relationship between the patient's age and the type of reporting medium preferred |
| Cuesta-Barriuso et al, 2018 [24] | • Medtep Hemophilia online platform (electronic diary and reminder)<br>• Documented infusion log, physical activities, and bleeding event<br>• Unrestricted real-time access by clinicians to patient's data<br>• Duration: 12 months | • Adherence to prophylactic treatment<br>• HRQoL<br>• Perception of illness<br>• Joint health | • Usage metrics | • Two-thirds of patients consistently had above 80% adherence.<br>• Significant increase in treatment adherence from baseline to 1 month, 6 months, and 12 months after intervention<br>• Significant improvement in HRQoL and illness perception from baseline to 12 months.<br>• No change in joint health |

[a]Not applicable.

[b]The primary objective of this study was to evaluate the efficacy of ReFacto, with the use of an electronic diary as the mode of documentation for bleeding events. Evaluating the effectiveness of electronic diary was an exploratory objective of this study.

[c]SMS: short message service.

[d]HRQoL: health-related quality of life.

# Discussion

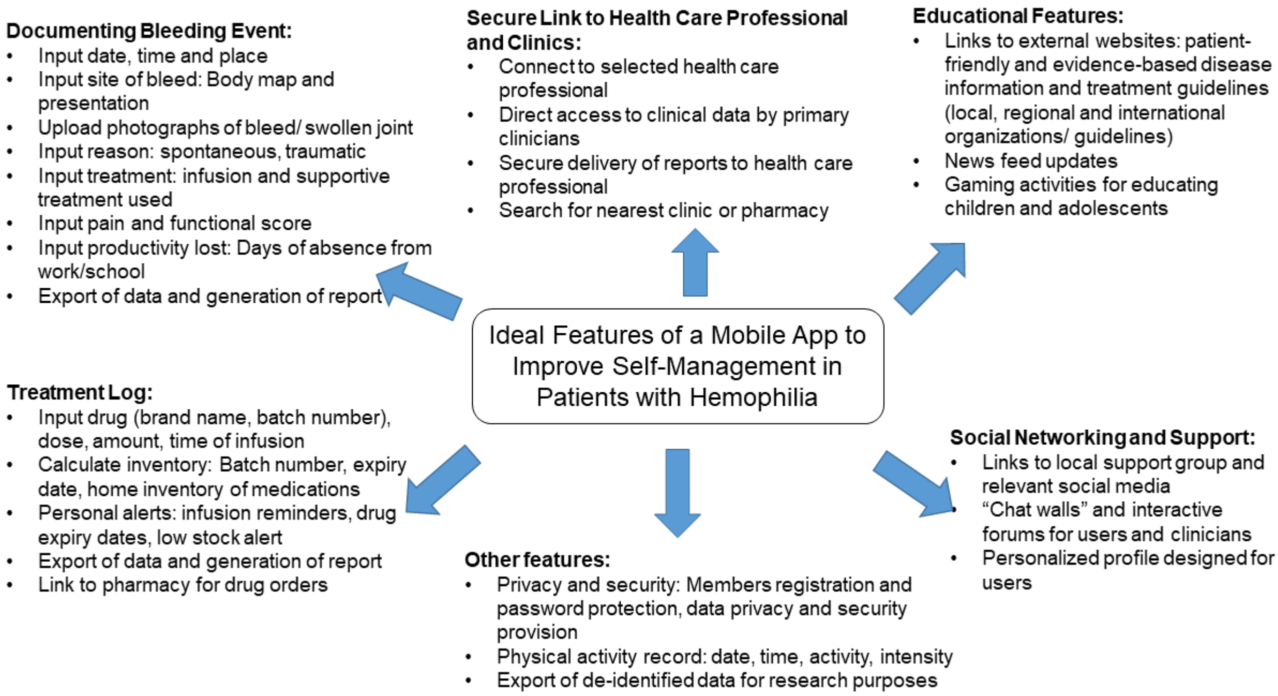
## Principal Findings

This systematic review summarizes the existing literature that evaluated the efficacy of distance-delivered technologies for improving adherence to prophylactic infusion and electronic recording of bleeding events and for ameliorating functional outcomes among patients with hemophilia. Specifically, there were too few high-quality studies from which to draw strong conclusions in support of the use of telehealth interventions in this population. Despite the need for more RCTs with larger samples to validate our findings, preliminary evidence supports the feasibility and effectiveness of interventions delivered by telehealth for improving self-management and health literacy among this population, especially among patients with severe hemophilia who require regular prophylactic infusions. However, the collective evidence seems to suggest that the technical errors and complex technological operations are major patient-related barriers. Additionally, the significant effects of telehealth interventions on symptom management were not consistently established.

Over the past decades, advances in treatment modalities have reduced the mortality and morbidity associated with hemophilia. However, the life-long maintenance of a demanding treatment regimen requires excellent self-management skills from both patients and their caregivers. Telehealth has become increasingly popular for providing integrated care to patients with chronic diseases. For patients with hemophilia, telemedicine-delivered integrated care includes the remote provision of education to improve self-efficacy. As good record keeping is an important aspect of home-based hemophilia care, our review also identified evidence to support the use of telemonitoring in order to enable increased adherence to the recording and transfer of clinical information, such as bleeding episodes and infusion logs. It is also worth noting that a telehealth intervention should not be administered alone. A handful of included studies adopted multimodal components to complement the telehealth technology, such as preintervention training and regular personal contact with the patient during the study period [12,14,16,22]. This multimodal approach may potentially enhance the user's proficiency, leading to increased acceptance and long-term adherence to the technology.

The growing popularity of mobile health (mHealth) over the past decade highlights the increasing trend involving the connection of patients with the world of digital health information via smartphones and other mobile devices [28]. One included study reported that that phone apps are associated with the most rapid reporting of bleeding episodes to the treatment center, with almost 40% of data being reported on the day of treatment and 70% reported within a week [13]. Based on the narrative syntheses in this review, we propose the features of an ideal mobile app for patients with hemophilia (Figure 2). This list comprises basic functions that allow the documentation of bleeding events and photographs, infusion logs, alerts, and a home inventory of medications. The comprehensive care of a patient with hemophilia also includes addressal of the patient's psychological needs, and in this regard, interactive platforms (eg, chat walls and forums) that foster a supportive social network within the hemophilia community would be an attractive feature, although caution has to be taken to ensure good netiquette among users. To promote sustained usage and adherence, one must also consider applying the principles of behavioral science when designing a mobile app from a technical perspective [29,30]. Applications that support personalization, allow user-friendly data recording, and provide well-paced reminders may more effectively promote behavioral changes.

**Figure 2.** Ideal features of a mobile app to improve self-management in patients with hemophilia.



Such an app should also adopt a multimedia approach to update disease- and treatment-related information. In this context, gamification deserves more attention and study, particularly as a means to engage adolescents who are becoming increasingly dependent in health management [31]. Reliable authentication systems must also be implemented to protect the patients' electronic health information, such as guidelines stipulated in the Health Insurance Portability and Accountability Act (HIPAA) in the United States or the Data Protection Act in the United Kingdom. Additionally, the ability to export de-identified data at the backend will provide a rich source of clinical data for research purposes.

### Clinical Implications and Direction for Future Research

Telehealth interventions can be used to provide quality health care to patients without readily accessible clinical services, such as those who reside in the inner city or rural areas. One study identified time, transportation difficulties, and the distance to the hematology clinic as the top barriers to obtaining care in the United States, especially for caregivers of children with hemophilia [32]. Beyond the reductions in travel costs, school-based telehealth programs may reduce the requirement for caregivers to secure time off from work and reduce the frequencies of emergency department and clinical visits. In developing countries, the lack of access to specialized clinics remains a major issue faced by patients who have recently initiated inhibitor therapy and require highly specific and closely monitored dosing. The introduction and systematic evaluation of telehealth programs in these areas could potentially expand patient access while reducing burdens such as travel required to receive specialty care, and improve monitoring, timeliness, and communication within the care continuum.

People with low levels of health literacy use more health care services, including visits to general practitioners, hospitals, and emergency care facilities [33,34]. This phenomenon may indicate that these patients bypass preventive care, adhere poorly to prophylaxis treatments, or are unable to make effective use of health care services [33]. A low level of health literacy is also associated with a reduced ability to take medications appropriately and a reduced adherence to chronic medication therapy. To address these issues, telehealth interventions can be implemented to promote motivation and ability among individuals by enabling access to understanding and processing of health-related information through improvements in cognitive and social skills. Built-in alerts and reminders within mobile apps may enhance adherence to prophylactic infusions. Of note, the three included studies involving online educational programs reported excellent outcomes in terms of improving patients' self-efficacy and theoretical knowledge regarding disease management [25-27]. Such interventions were designed to be more engaging than traditional patient information leaflets or booklets by using more audiovisual information and interactions intended to demolish the literacy barrier. As hemophilia is a chronic disease characterized by complex care needs, an appealing telehealth-delivered educational intervention may play an important role in empowering adolescents as they transition to adulthood and take a more independent role in managing their health.

Cultural factors must be considered when selecting the most appropriate delivery of technology. Research has suggested that the levels of patient engagement and health literacy differ by race and ethnicity. Distance-delivered interventions for patients with hemophilia should include culturally tailored patient education programs and materials. For example, one qualitative study discussed the subjective illness experiences of patients with hemophilia in the United States and United Kingdom, with a particular focus on cultural and social contexts [35]. The authors found that patients in the United States tended to more strongly emphasize on existing support systems, such as relationships with health care practitioners or the cost of health insurance, whereas patients in the United Kingdom considered functional problems related to pain and disability more relevant in their everyday lives [35]. In addition to considering the involvement of cultural context in every aspect of the care continuum, the development of a telehealth intervention should also focus on the patients' personal experiences of sickness and its effects on other relational contexts, such as family, school, or work. Most commercially available social networking apps currently used for patients with hemophilia were developed using English as the primary language. Therefore, future research should evaluate the needs of patients in non-English–speaking countries and thus elucidate the role of sociocultural variables in modifying the experience of this disease through advanced technology.

Telehealth interventions can potentially facilitate regional and international "teletwinning" to allow collaborative research and harmonization of data collection from patients with hemophilia [4]. One included study used a handheld computer diary platform to gather data on bleeding episodes from patients with severe or moderate hemophilia A in four Nordic countries (Denmark, Finland, Norway, and Sweden) [17]. The authors concluded that this platform not only improved adherence to reporting but also enabled the long-term postmarketing surveillance of treatment products by pooling data from different regions. In recent years, mHealth has emerged to augment specialized health care services delivered to underserved populations, especially in countries with high levels of health disparity such as India [36] and mainland China [37]. Teletwinning between developed and developing countries may bridge health care disparities and extend the highest standards of hemophilia care to all patients.

In addition to evaluating the efficacy, it is important to consider the cost-effectiveness of telehealth interventions for hemophilia management. Only one included study reported the costs associated with administering an SMS intervention from the economic perspectives of the participant and institution [21]. Compared to the traditional modes of health care delivery, new technologies will inevitably demand resources to develop, customize, initiate, and maintain the platforms and ensure the efficient achievement of the intended purposes. Data on the value and cost-effectiveness of such technologies would be indispensable if the eventual goal is to implement telehealth interventions in public health care systems on a wide scale. However, we acknowledge that such comparative studies may be methodologically difficult to conduct, as the hemophilia population is small and would require quantitative outcomes such as mortality and emergency department visits due to bleeding episodes, collected over a longer time horizon.

### Limitations

Hemophilia is a rare disease, and this review was limited by the inclusion of studies with a small sample size and lacking a control group. The wide variability in outcome measures and intervention regimens as well as the reliance on self-reported efficacy and satisfaction measures limited our ability to compare and draw conclusive results regarding the effects of the interventions. Heterogeneity in the assessment of health outcomes across studies also made it difficult to determine the effectiveness of the health technologies. Lastly, we acknowledge that even though a protocol was developed prior to the initiation of this review, it was not prospectively registered with any registries that facilitate public scrutiny. However, much effort has been dedicated to ensure strict adherence to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and transparency and comprehensiveness in reporting. Given these limitations, our collective findings should be interpreted cautiously.

### Conclusions

To our knowledge, this is the first systematic review to evaluate the collective evidence of telehealth-delivered interventions for patients with hemophilia. Most interventions involved reminding patients to administer prophylactic infusions and documenting bleeding events and treatment logs. Although little explicit evidence is available, telehealth-delivered interventions could feasibly improve patients' adherence to medication use and promote independence in disease management. Multimedia educational platforms appear to effectively enhance knowledge transfer and information utilization, particularly among adolescent and young adult patients. However, the sustainability of use and long-term adherence to the intervention remain uncertain. Given the complexity and resources involved in developing a mature and established system, support from a dedicated network of hemophilia specialists and data managers will be required to maintain the technology, improve adherence to prophylactic treatment and recording, and validate the electronic data locally. Future research should also involve RCTs with longer time horizons to investigate the effects of interventions delivered by telehealth on improved health outcomes and behaviors (eg, physical activity) among patients with hemophilia.

## Conflicts of Interest

None declared.

## References

1. Chen S. Economic costs of hemophilia and the impact of prophylactic treatment on patient management. Am J Manag Care 2016 Apr;22(5 Suppl):s126-s133 [FREE Full text] [Medline: 27266809]
2. Health Resources & Services Administration. 2018. Telehealth Programs URL: https://www.hrsa.gov/rural-health/telehealth/index.html [accessed 2019-06-21]
3. Larsen SB, Sørensen NS, Petersen MG, Kjeldsen GF. Towards a shared service centre for telemedicine: Telemedicine in Denmark, and a possible way forward. Health Informatics J 2016 Dec;22(4):815-827 [FREE Full text] [doi: 10.1177/1460458215592042] [Medline: 26261216]
4. Kulkarni R. Use of telehealth in the delivery of comprehensive care for patients with haemophilia and other inherited bleeding disorders. Haemophilia 2018 Jan;24(1):33-42. [doi: 10.1111/hae.13364] [Medline: 29205699]
5. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer. J Med Internet Res 2017 Dec 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
6. Hernandez C, Mallow J, Narsavage G. Delivering telemedicine interventions in chronic respiratory disease. Breathe (Sheff) 2014 Sep;10(3):198-212 [FREE Full text] [doi: 10.1183/20734735.008314] [Medline: 26843894]
7. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
8. Tricco A, Lillie E, Zarin W, O'Brien K, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
9. McGowan P. Self-management education and support in chronic disease management. Prim Care 2012 Jun;39(2):307-325. [doi: 10.1016/j.pop.2012.03.005] [Medline: 22608868]
10. Armijo-Olivo S, Stiles C, Hagen N, Biondo P, Cummings G. 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]
11. Viera A, Garrett J. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]

12. Collins P, Bolton-Maggs P, Stephenson D, Jenkins B, Loran C, Winter M. Pilot study of an Internet-based electronic patient treatment record and communication system for haemophilia, Advoy.com. Haemophilia 2003 May;9(3):285-291. [Medline: 12694519]
13. Hay C, Xiang H, Scott M, Collins P, Liesner R, Dolan G, et al. The haemtrack home therapy reporting system: Design, implementation, strengths and weaknesses: A report from UK Haemophilia Centre Doctors Organisation. Haemophilia 2017 Sep;23(5):728-735. [doi: 10.1111/hae.13287] [Medline: 28806858]
14. Walker I, Sigouin C, Sek J, Almonte T, Carruthers J, Chan A, et al. Comparing hand-held computers and paper diaries for haemophilia home therapy: a randomized trial. Haemophilia 2004 Nov;10(6):698-704. [doi: 10.1111/j.1365-2516.2004.01046.x] [Medline: 15569164]
15. Arnold E, Heddle N, Lane S, Sek J, Almonte T, Walker I. Handheld computers and paper diaries for documenting the use of factor concentrates used in haemophilia home therapy: a qualitative study. Haemophilia 2005 May;11(3):216-226. [doi: 10.1111/j.1365-2516.2005.01095.x] [Medline: 15876266]
16. Pattacini C, Rivolta G, Di Perna C, Riccardi F, Tagliaferri A, Haemophilia Centres Network of Emilia-Romagna Region. A web-based clinical record 'xl'Emofilia' for outpatients with haemophilia and allied disorders in the Region of Emilia-Romagna: features and pilot use. Haemophilia 2009 Jan;15(1):150-158. [doi: 10.1111/j.1365-2516.2008.01921.x] [Medline: 18976246]
17. Petrini P, Rylander C. Clinical safety surveillance study of the safety and efficacy of long-term home treatment with ReFacto utilizing a computer-aided diary: a Nordic multicentre study. Haemophilia 2009 Jan;15(1):175-183. [doi: 10.1111/j.1365-2516.2008.01863.x] [Medline: 18752534]
18. Mondorf W, Siegmund B, Mahnel R, Richter H, Westfeld M, Galler A, et al. Haemoassist--a hand-held electronic patient diary for haemophilia home care. Haemophilia 2009 Mar;15(2):464-472. [doi: 10.1111/j.1365-2516.2008.01941.x] [Medline: 19226411]
19. Vallée-Smejda S, Hahn M, Aubin N, Rosmus C. Recording practices and satisfaction of hemophiliac patients using two different data entry systems. Comput Inform Nurs 2009;27(6):372-378. [doi: 10.1097/NCN.0b013e3181bcad12] [Medline: 19901574]
20. Leone J. Utility of a wireless, handheld monitoring system in the management of hemophilia patients. Comput Inform Nurs 2011 Sep;29(9):521-522. [doi: 10.1097/NCN.0b013e3182066356] [Medline: 21239988]
21. Broderick C, Herbert R, Latimer J, Mathieu E, van Doorn N, Curtin J. Feasibility of short message service to document bleeding episodes in children with haemophilia. Haemophilia 2012 Nov;18(6):906-910. [doi: 10.1111/j.1365-2516.2012.02869.x] [Medline: 22681182]
22. Young G, Solem C, Hoffman K, Kabawat J, Pickard A, Gut R, et al. Capturing daily assessments and home treatment of congenital hemophilia with inhibitors: design, disposition, and implications of the Dosing Observational Study in Hemophilia (DOSE). J Blood Med 2012;3:131-138 [FREE Full text] [doi: 10.2147/JBM.S37016] [Medline: 23152717]
23. Sholapur N, Barty R, Wang G, Almonte T, Heddle N. A survey of patients with haemophilia to understand how they track product used at home. Haemophilia 2013 Sep;19(5):e289-e295. [doi: 10.1111/hae.12170] [Medline: 23672744]
24. Cuesta-Barriuso R, López-Pina JA, Nieto-Munuera J, Sagarra-Valls G, Panisello-Royo J, Torres-Ortuño A. Effectiveness of the Medtep Hemophilia online platform for adherence to prophylactic treatment in haemophilia patients: Results from a 1-year observational study. Haemophilia 2018 May;24(3):452-459. [doi: 10.1111/hae.13444] [Medline: 29493864]
25. Mulders G, de Wee E, Vahedi Nikbakht-Van de Sande M, Kruip M, Elfrink E, Leebeek F. E-learning improves knowledge and practical skills in haemophilia patients on home treatment: a randomized controlled trial. Haemophilia 2012 Sep;18(5):693-698. [doi: 10.1111/j.1365-2516.2012.02786.x] [Medline: 22458978]
26. Breakey V, Ignas D, Warias A, White M, Blanchette V, Stinson J. A pilot randomized control trial to evaluate the feasibility of an Internet-based self-management and transitional care program for youth with haemophilia. Haemophilia 2014 Nov;20(6):784-793. [doi: 10.1111/hae.12488] [Medline: 25311370]
27. Breakey V, Warias A, Ignas D, White M, Blanchette V, Stinson J. The value of usability testing for Internet-based adolescent self-management interventions: "Managing Hemophilia Online". BMC Med Inform Decis Mak 2013 Oct 04;13:113 [FREE Full text] [doi: 10.1186/1472-6947-13-113] [Medline: 24094082]
28. Marcolino M, Oliveira J, D'Agostino M, Ribeiro A, Alkmim M, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
29. Stroulia E, Fairbairn S, Bazelli B. IEEE. 2013. Smart-phone application design for lasting behavioral changes URL: https://ieeexplore.ieee.org/document/6627804 [accessed 2019-06-24]
30. Mendiola M, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
31. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]

32. Zhou Z, Riske B, Forsberg A, Ullman M, Baker J, Koerper M, et al. Self-reported barriers to hemophilia care in people with factor VIII deficiency. Am J Prev Med 2011 Dec;41(6 Suppl 4):S346-S353. [doi: 10.1016/j.amepre.2011.09.003] [Medline: 22099357]
33. Berkman N, Sheridan S, Donahue K, Halpern D, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
34. Griffey R, Kennedy S, D'Agostino McGowan L, McGownan L, Goodman M, Kaphingst KA. Is low health literacy associated with increased emergency department utilization and recidivism? Acad Emerg Med 2014 Oct;21(10):1109-1115 [FREE Full text] [doi: 10.1111/acem.12476] [Medline: 25308133]
35. Palareti L, Potì S, Cassis F, Emiliani F, Matino D, Iorio A. Shared topics on the experience of people with haemophilia living in the UK and the USA and the influence of individual and contextual variables: Results from the HERO qualitative study. Int J Qual Stud Health Well-being 2015;10:28915 [FREE Full text] [doi: 10.3402/qhw.v10.28915] [Medline: 26578360]
36. Bhan N, Rao KD, Kachwaha S. Health inequalities research in India: a review of trends and themes in the literature since the 1990s. Int J Equity Health 2016 Dec 06;15(1):166-123 [FREE Full text] [doi: 10.1186/s12939-016-0457-y] [Medline: 27716296]
37. Hong YA, Zhou Z, Fang Y, Shi L. The Digital Divide and Health Disparities in China: Evidence From a National Survey and Policy Implications. J Med Internet Res 2017 Dec 11;19(9):e317 [FREE Full text] [doi: 10.2196/jmir.7786] [Medline: 28893724]

## Abbreviations

**EPHPP:** Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies
**Hem:** hemophilia
**HemA:** hemophilia A
**HemB:** hemophilia B
**HIPAA:** Health Insurance Portability and Accountability Act
**HRQoL:** health-related quality of life
**IQR:** interquartile range
**ITT:** immune tolerance treatment
**NR:** not reported
**OS:** observational study
**RCT:** randomized controlled trial
**SMS:** short message service

*Edited by G Eysenbach; submitted 08.10.18; peer-reviewed by S Prior, R Ramsey, R Grainger; comments to author 01.04.19; revised version received 14.05.19; accepted 03.06.19; published 10.07.19.*

*Please cite as:*
*Qian W, Lam TTN, Lam HHW, Li CK, Cheung YT*
*Telehealth Interventions for Improving Self-Management in Patients With Hemophilia: Scoping Review of Clinical Studies*
*J Med Internet Res 2019;21(7):e12340*
*URL: http://www.jmir.org/2019/7/e12340/*
*doi:10.2196/12340*
*PMID:31293241*

Original Paper

# Association of Remote Monitoring With Survival in Heart Failure Patients Undergoing Cardiac Resynchronization Therapy: Retrospective Observational Study

Peter Bogyi[1,2], MD; Mate Vamos[3], MD, PhD; Zsolt Bari[1], MD; Balazs Polgar[1], MD; Balazs Muk[1], MD; Noemi Nyolczas[1], MD, PhD; Robert Gabor Kiss[1], MD, PhD; Gabor Zoltan Duray[1,2], MD, PhD

[1]Department of Cardiology, Hungarian Defence Forces Medical Centre, Budapest, Hungary

[2]Basic and Translational Medicine, Karoly Racz School of PhD Studies, Semmelweis University, Budapest, Hungary

[3]Department of Cardiology, University Hospital Frankfurt, Goethe University, Frankfurt, Germany

**Corresponding Author:**
Gabor Zoltan Duray, MD, PhD
Department of Cardiology
Hungarian Defence Forces Medical Centre
44 Robert Karoly Avenue
Budapest, 1134
Hungary
Phone: 36 14651800 ext 71759
Fax: 36 14651857
Email: gduray@yahoo.com

## Abstract

**Background:** Remote monitoring is an established, guideline-recommended technology with unequivocal clinical benefits; however, its ability to improve survival is contradictory.

**Objective:** The aim of our study was to investigate the effects of remote monitoring on mortality in an optimally treated heart failure patient population undergoing cardiac resynchronization defibrillator therapy (CRT-D) implantation in a large-volume tertiary referral center.

**Methods:** The population of this single-center, retrospective, observational study included 231 consecutive patients receiving CRT-D devices in the Medical Centre of the Hungarian Defence Forces (Budapest, Hungary) from January 2011 to June 2016. Clinical outcomes were compared between patients on remote monitoring and conventional follow-up.

**Results:** The mean follow-up time was 28.4 (SD 18.1) months. Patients on remote monitoring were more likely to have atrial fibrillation, received heart failure management at our dedicated heart failure outpatient clinic more often, and have a slightly lower functional capacity. Crude all-cause mortality of remote-monitored patients was significantly lower compared with patients followed conventionally (hazard ratio [HR] 0.368, 95% CI 0.186-0.727, $P$=.004). The survival benefit remained statistically significant after adjustment for important baseline parameters (adjusted HR 0.361, 95% CI 0.181-0.722, $P$=.004).

**Conclusions:** In this single-center, retrospective study of optimally treated heart failure patients undergoing CRT-D implantation, the use of remote monitoring systems was associated with a significantly better survival rate.

***(J Med Internet Res 2019;21(7):e14142)*** doi:10.2196/14142

**KEYWORDS**

survival; CRT-D; remote monitoring; telemedicine; heart failure

## Introduction

Remote monitoring of cardiac implantable electronic devices has proved to be beneficial on several clinical endpoints. The first studies confirmed the feasibility of early, automatic detection of technical issues [1], and recognition of a new onset of atrial fibrillation [2]. Randomized studies proved that remote monitoring could reduce time to evaluate arrhythmic events [3], decrease mean length of cardiovascular hospitalizations [4], and could significantly lower the number of appropriate or inappropriate shocks [5]. This method is also able to reduce in-office implantable cardioverter defibrillator (ICD) follow-up burden safely [6] Remote monitoring also provided early detection of heart failure events and reduced the number of urgent in-office visits and total health care use in patients with ICD or cardiac resynchronization systems in diverse clinical studies [7-9].

Moreover, registry data suggest a potential survival benefit in patients on remote monitoring [10,11]. The most important limitation of these register-based reports is the lack of randomization and the paucity of clinical characteristics that would make a more accurate comparison possible. In the multicenter EFFECT study, remote monitoring was associated with reduced deaths and cardiovascular hospitalizations in patients with ICD. [12] In the randomized, controlled, international multicenter IN-TIME study, a significant survival benefit of implant-based multiparameter telemonitoring was demonstrated over the standard of care in patients with heart failure and implanted dual-chamber ICD or cardiac resynchronization therapy defibrillator (CRT-D) [13].

However, a recent randomized trial on remote monitoring (MORE-CARE) could not reduce mortality or risk of cardiovascular or device-related hospitalizations [14]. Furthermore, the REM-HF multicenter randomized study showed similar outcomes among patients with heart failure and cardiac implantable electronic devices utilizing remote monitoring with weekly downloads and a prespecified follow-up approach [15].

Concerning these contradictory results, we aimed to investigate the effects of remote monitoring on mortality in an optimally treated heart failure patient population undergoing CRT-D implantation in a large-volume tertiary referral center.

## Methods

### Study Patients

The population of this single-center, retrospective, observational study included consecutive patients receiving CRT-D devices in the Medical Centre of the Hungarian Defence Forces (Budapest, Hungary) from January 2011 to June 2016. Indication for implantation was established according to current guideline recommendations of the European Society of Cardiology [16,17]. CRT-ICDs from various manufacturers were used (Biotronik, Germany; ELA/Sorin, Italy; Guidant/Boston Scientific, Marlborough, MA; Medtronic, Minneapolis, MN; and St Jude Medical/Abbott, St Paul, MN). Choice of device type was left to the implanting physician's discretion.

### Study Design

The possibility of remote monitoring was offered to every patient implanted with a wireless telemetry-capable CRT-D device. Patients who consented to remote monitoring formed the remote monitoring group. During the inclusion period, remote monitoring systems of two manufacturers (Medtronic CareLink Network, Medtronic Inc, Minneapolis, MN; and Home Monitoring Service Center, Biotronik GmbH & Co KG, Germany) were available in our center. The control group consisted of CRT-D recipients, who were followed up in our outpatient device clinic without remote monitoring.

The CareLink network operates with scheduled transmissions defined by the physician, and unscheduled transmissions, which can be triggered both by the patient (manual transmission) and by the device itself (alert event). Scheduled remote transmissions were set up every 3 months [18]. It was also recommended to patients to send manual transmissions in case of palpitation, syncope, or worsening of heart failure symptoms. Alert programming was set up according to previously published Medtronic-sponsored trials as follows: OptiVol alert (nominal fluid index ≥60 Ω-day), daily atrial fibrillation burden greater than 6 hours per day, ventricular rate during atrial fibrillation greater than 100 bpm for 6 hours, two or more shocks delivered, all therapies exhausted, lead or device integrity alert, lead impedance out of range, recommended replacement time, and end of service [7,18,19].

Home Monitoring uses a mobile phone network to transmit device data automatically on a daily basis, as well as instantly on the occurrence of a potentially clinically relevant event [20]. These parameters include device and battery status, pacing impedances, bradycardia, tachycardia, and CRT statistics, mean heart rates, patient activity, heart rate variability, and current programming of the device [21].

In-office visits were recommended to patients on remote monitoring without symptoms at least once a year [22]. Remote transmissions were evaluated every day by a team consisting of cardiology trainees and consultant electrophysiologists. In the case of suspected heart failure progression, a heart failure specialist was involved additionally. Transmissions were labeled as clinically nonsignificant, clinically relevant (yellow alert), or highly urgent (red alert). If a clinically relevant event was perceived, patients were contacted via phone calls and were invited to the clinic for a personal visit in a week. Definitions of a clinically relevant event for the two remote monitoring systems are summarized in Multimedia Appendix 1. Transmissions with highly urgent content defined as ventricular arrhythmias treated with more than one ICD shock or system integrity alert were handled within 24 hours. Patients with missed transmissions longer than 4 weeks were also contacted. Follow-up of patients in the control group was performed at intervals of 3 to 6 months.

Patient demographics, comorbidities, pharmacotherapy, electrocardiogram characteristics, echocardiography, and laboratory data were collected at enrollment and during scheduled remote checks or in-office visits.

All patients on remote monitoring signed a related informed consent form. The study complied with the Declaration of Helsinki, and the study protocol was approved by the Institutional Ethics Committee of Hungarian Defence Forces Medical Centre, Budapest, Hungary.

### Study Endpoints

The primary objective of this study was to compare the mortality of remote-monitored patients with patients on conventional follow-up. Survival was assessed as the time from CRT-D implantation to all-cause mortality. Mortality data were retrieved using the Hungarian National Health Fund Death Registry. The unique health insurance number of a patient is deactivated immediately after death. The secondary endpoint was the response to resynchronization therapy at the visit at 6 to 12 months, defined as 5% absolute increase in left ventricular ejection fraction. The number of all ambulant visits, device clinic visits, and heart failure outpatient clinic visits were also analyzed and compared between the two patient groups.

### Statistical Analysis

Statistical analysis was performed using PASW Statistics software, version 18.0.0 (WinWrap Basic, Polar Engineering and Consulting). The Kolmogorov-Smirnov test was used to evaluate the normal distribution of continuous data. The chi-square test was applied to test for categorical variables; the two-sample *t* test or the Mann-Whitney *U* test was used for continuous variables among patient groups.

To assess the effects of remote monitoring on survival, the Cox proportional hazards regression model was used. Univariate analysis was performed for the following variables: age; gender; heart failure management; upgrade procedure; secondary prevention; ischemic etiology; atrial fibrillation; hypertension; hyperlipidemia; diabetes mellitus; stroke; peripheral artery disease; chronic obstructive pulmonary disease; New York Heart Association (NYHA) functional class; left ventricular ejection fraction; QRS duration; left bundle branch block; estimated glomerular filtration rate; hemoglobin; and therapy with platelet aggregation inhibitors, beta blockers, angiotensin-converting-enzyme inhibitors or angiotensin-receptor blockers, mineralocorticoid antagonists, diuretics, digitalis, amiodarone, and statin. All variables with $P \leq .10$ on univariate analysis were included in the multivariate Cox models. Two-sided $P$ values <.05 were considered statistically significant. Survival curves were constructed according to the Kaplan-Meier method and compared with the Cox proportional hazard model. To check for interaction between survival and the specific remote monitoring system, all-cause mortality was also compared between the subgroups of patients on CareLink and on Home Monitoring systems.

## Results

A total of 231 CRT-D recipients were included in this study. Of the 90 patients implanted with remote monitoring-capable devices, 62 consented to receive a remote monitor (41 of 56 patients with Medtronic and 21 of 34 with Biotronik devices; Figure 1).

Detailed patient baseline data are summarized in Table 1. Despite the nonrandomized nature of the study, there were only a few significant differences between the two patient groups: patients on remote monitoring were more likely to have atrial fibrillation and have received heart failure management more often at our dedicated heart failure outpatient clinic. They also had a slightly lower NYHA functional class.

**Figure 1.** Study flow diagram. CRT-D: cardiac resynchronization therapy defibrillator; RM: remote monitoring.

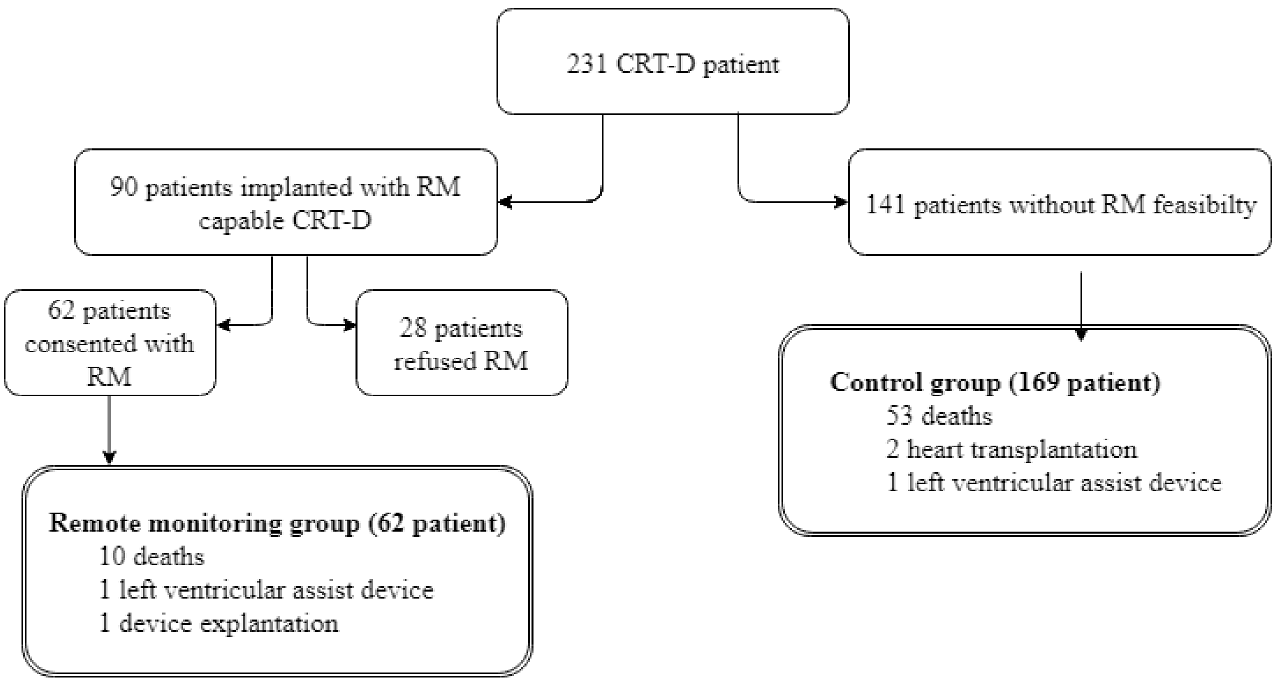


**Table 1.** Patient baseline characteristics (N=231).

| Characteristic | Remote monitoring group (n=62) | Control group (n=169) | *P* value |
|---|---|---|---|
| Age (years), mean (SD) | 64.0 (9.9) | 65.6 (10.8) | .26 |
| Male, n (%) | 51 (82) | 133 (78.7) | .55 |
| Left ventricular ejection fraction (%), mean (SD) | 27.3 (6.5) | 26.3 (5.8) | .29 |
| NYHA[a] functional class, mean (SD) | 2.4 (0.7) | 2.6 (0.7) | .05 |
| QRS duration (ms), mean (SD) | 147 (28) | 154 (29) | .08 |
| Left bundle branch block, n (%) | 46 (74) | 130 (76.9) | .71 |
| Upgrade procedure, n (%) | 10 (16) | 46 (27.2) | .08 |
| Primer prevention, n (%) | 51 (82) | 136 (80.5) | .76 |
| Heart failure outpatient clinic management, n (%) | 39 (63) | 77 (45.6) | .02 |
| Ischemic etiology, n (%) | 33 (53) | 95 (56.2) | .68 |
| Hypertension, n (%) | 46 (74) | 131 (77.5) | .14 |
| Diabetes, n (%) | 25 (40) | 54 (31.9) | .23 |
| Hyperlipidemia, n (%) | 25 (40) | 51 (30.2) | .14 |
| Atrial fibrillation, n (%) | 25 (40) | 44 (26.0) | .03 |
| Chronic obstructive pulmonary disease, n (%) | 9 (15) | 22 (13) | .76 |
| Estimated glomerular filtration rate (mL/min/1.73 $m^2$), mean (SD) | 55.9 (17.6) | 58.3 (20.3) | .58 |
| Hemoglobin (g/L), mean (SD) | 133 (15) | 133 (16) | .95 |
| **Concomitant medications, n (%)** | | | |
| Beta blocker | 62 (100) | 166 (98.2) | .29 |
| ACEi[b]/ARB[c] | 61 (98) | 167 (98.8) | .89 |
| Mineralocorticoid receptor antagonist | 58 (94) | 158 (93.5) | .98 |
| Diuretic | 56 (90) | 157 (92.9) | .51 |
| Amiodarone | 14 (23) | 51 (30.2) | .25 |
| Digoxin | 16 (26) | 26 (15.4) | .06 |

[a]NYHA: New York Heart Failure Association.

[b]ACEi: angiotensin-converting-enzyme inhibitor.

[c]ARB: angiotensin-receptor blocker.

During the average follow-up time of 28.4 (SD 18.1) months, 63 patients died, 2 underwent heart transplantation, 2 received a left ventricular assist device, and in 1 case device explantation was performed due to infection. Crude all-cause mortality of remote-monitored patients was significantly lower compared with patients followed conventionally (hazard ratio [HR] 0.368, 95% CI 0.186-0.727, *P*=.004; Table 2; Figure 2). The survival benefit remained statistically significant after adjustment for important baseline parameters (adjusted HR 0.361, 95% CI 0.181-0.722, *P*=.004; Table 3; Figure 2). The survival benefit did not differ between the remote monitoring systems (ie, CareLink vs Home Monitoring; *P*=.79).

Echocardiographic response to cardiac resynchronization therapy at 6 to 12 months was more often observed in patients on remote monitoring (41.9%, 26/62 vs 31.9%, 54/169); however, this difference was statistically nonsignificant (Table 4).

**Table 2.** Predictors of mortality (univariate Cox regression).

| Predictor | Hazard ratio (95% CI) | *P* value |
|---|---|---|
| Age | 1.049 (1.020-1.079) | .001 |
| Male gender | 1.996 (0.949-4.195) | .07 |
| Remote monitoring | 0.368 (0.186-0.727) | .004 |
| Heart failure management | 0.584 (0.351-0.970) | .04 |
| Upgrade | 1.869 (1.112-3.140) | .02 |
| Secondary prevention | 1.488 (0.820-2.698) | .19 |
| Ischemic etiology | 1.373 (0.285-2.284) | .22 |
| Atrial fibrillation | 1.799 (1.083-2.990) | .02 |
| Hypertension | 2.226 (1.128-4.395) | .02 |
| Hyperlipidemia | 0.811 (0.469-1.402) | .45 |
| Diabetes mellitus | 1.092 (0.650-1.836) | .74 |
| Stroke | 1.699 (0.861-3.354) | .13 |
| Peripheral artery disease | 1.155 (0.663-3.668) | .31 |
| Chronic obstructive pulmonary disease | 1.239 (0.587-2.616) | .57 |
| Platelet aggregation inhibitor | 1.180 (0.714-1.948) | .52 |
| Beta blocker | 0.407 (0.099-1.684) | .22 |
| ACEi[a]/ARB[b] | 0.692 (0.096-5.007) | .72 |
| Mineralocorticoid antagonist | 0.769 (0.349-1.696) | .52 |
| Diuretics | 1.606 (0.503-5.217) | .42 |
| Digitalis | 0.772 (0.410-1.454) | .42 |
| Amiodaron | 1.986 (1.192-3.309) | .008 |
| Statin | 1.243 (0.725-2.130) | .43 |
| NYHA[c] functional class | 1.394 (1.017-1.910) | .04 |
| Left ventricular ejection fraction | 1.028 (0.987-1.071) | .18 |
| QRS duration | 1.001 (0.992-1.010) | .88 |
| Left bundle branch block | 1.308 (0.961-1.782) | .09 |
| Estimated glomerular filtration rate | 0.991 (0.978-1.005) | .23 |
| Hemoglobin | 1.002 (0.998-1.007) | .26 |

[a]ACEi: angiotensin-converting-enzyme inhibitor.

[b]ARB: angiotensin-receptor blocker.

[c]NYHA: New York Heart Failure Association.

**Figure 2.** Kaplan-Meier curves for all-cause mortality by follow-up type (remote monitoring vs conventional follow-up). HR: hazard ratio.

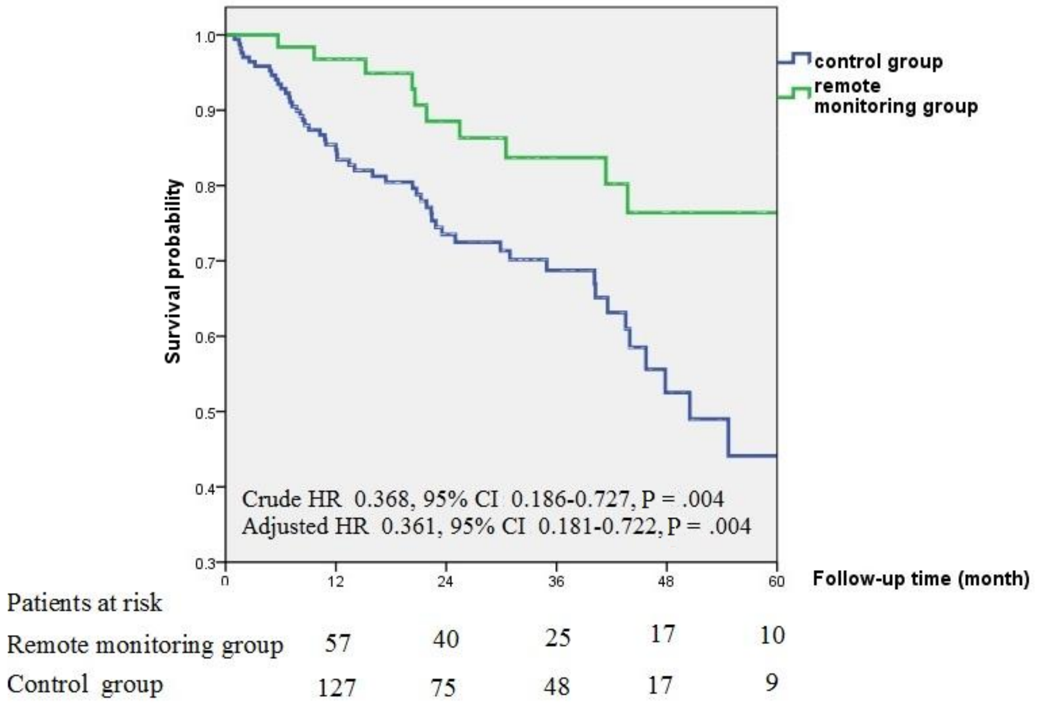



**Table 3.** Independent predictors of mortality (multivariate Cox regression).[a]

| Predictor | Hazard ratio (95% CI) | *P* value |
|---|---|---|
| Age | 1.035 (1.007-1.065) | .02 |
| Remote monitoring | 0.361 (0.181-0.722) | .004 |
| Amiodaron | 1.732 (1.032-2.907) | .04 |
| Atrial fibrillation | 1.727 (1.019-2.926) | .04 |

[a]Cases with missing values 0%.

**Table 4.** Echocardiographic response to cardiac resynchronization therapy at 6 to 12 months (N=231).

| Left ventricular ejection fraction (%) | Remote monitoring group (n=62) | Control group (n=169) | *P* value |
|---|---|---|---|
| Baseline, mean (SD) | 27.3 (6.5) | 26.3 (5.8) | .29 |
| Follow-up, mean (SD) | 33.2 (8.2) | 32.3 (9.2) | .34 |
| Change, mean (SD) | 6.9 (8.0) | 6.2 (8.2) | .48 |
| Absolute improvement ≥5%, n (%) | 26 (41.9) | 54 (31.9) | .57 |

The total number of follow-up controls tended to be higher in the remote monitoring group compared with patients undergoing conventional follow-up (4.6 visits per patient per year vs 3.9 visits per patient per year, *P*=.08). This was because patients on remote monitoring presented at our specialized heart failure outpatient clinic more often (1.9 visits per patient per year vs 1.1 visits per patient per year, *P*=.003; Table 5).

Of the 41 patients followed with the CareLink system, 1083 transmissions were received during the follow-up period. Seven transmissions contained highly urgent clinical information (four appropriate shock episodes and three lead integrity alerts). Moreover, two patients reached an elective replacement indicator. In addition, 396 transmissions included OptiVol alerts. Telemonitoring observations in the 21 patients on the Biotronik Home Monitoring system were as follows: 3 red alerts (two electric storms, one end of service) and 85 yellow alerts (8 sustained ventricular arrhythmias requiring antitachycardia pacing or shock therapy, 36 supraventricular tachyarrhythmia, 36 low biventricular pacing percentage, and 5 elevated left ventricular threshold).

**Table 5.** Ambulatory visits during follow-up (N=231).

| Visit type | Remote monitoring group (n=62) | Control group (n=169) | *P* value |
|---|---|---|---|
| **All ambulant visits** | | | .08 |
| Visits, n | 711 | 1187 | |
| Visits per patient per year, mean (SD) | 4.6 (3.0) | 3.9 (3.0) | |
| **Device clinic visits** | | | .95 |
| Visits, n | 435 | 889 | |
| Visits per patient per year, mean (SD) | 3.1 (2.4) | 2.9 (2.5) | |
| **Heart failure outpatient clinic visits** | | | .003 |
| Visits, n | 347 | 344 | |
| Visits per patient per year, mean (SD) | 1.9 (2.4) | 1.1 (2.0) | |

# Discussion

## Principal Findings

In this single-center, retrospective, observational study of 231 CRT-D recipients, use of remote monitoring was associated with better survival compared with patients undergoing conventional follow-up. The total number of follow-up visits was not reduced with this technique in our cohort.

## Comparison With Prior Work

Remote monitoring systems proved to be feasible, reliable, accessible, and were supported by the current guideline recommendations [16,23]. They are still underutilized despite the clear advantages, such as early identification of device malfunction or arrhythmic events. Regarding survival benefit, the available clinical data are controversial [10,13,24-27].

There are several proposed mechanisms contributing to the improved clinical outcome: early detection of clinically relevant arrhythmias (ie, atrial fibrillation or ventricular tachycardia) and early recognition of device malfunctions or suboptimal programming, which avoids unnecessary shocks and achieves proper percentage of biventricular pacing, respectively.

Moreover, the number of OptiVol alerts and the related visits at the heart failure outpatient clinic suggest that the observed mortality benefit associated with remote monitoring was also driven by early response to cardiac decompensation in our study. OptiVol is a detection algorithm developed for early recognition of cardiac decompensation using changes of intrathoracic impedance as a marker of lung fluid status [28]. This method, used for the detection of cardiac decompensation, is considered to be a very sensitive but less specific tool, which leads to a high number of false positive alerts. We have previously described a refined device diagnostic algorithm based on parameters of low activity level, high nocturnal heart rate, and suboptimal biventricular pacing, which could significantly improve clinical reliability of OptiVol alerts [9]. However, a recent OptiLink heart failure study analyzing this technology failed to demonstrate any difference in the composite endpoint of all-cause mortality and cardiovascular hospitalization [29]. Notably, only 30.3% of intrathoracic fluid index threshold crossing led to medical action, and it led to altered medication in only 26.0% of patients.

Our analysis shows a survival benefit for patients on remote monitoring; therefore, we are convinced that remote monitoring should be part of the follow-up of patients with CRT-D devices. One possible reason for the conflicting results of previous studies is the wide variety of actions on remote monitoring findings. A recently published meta-analysis by Stockburger et al [30] showed that device-based remote monitoring strategies specifying close-meshed comprehensive data acquisition and defined treatment interventions are able to significantly reduce total mortality and cardiovascular mortality, whereas remote data acquisition alone without specified treatment interventions appears to be ineffective on hard endpoints. The significantly increased number of visits in the heart failure clinic in our study supports this hypothesis, as the demonstrated survival benefit could have been achieved by more frequent follow-up and focused treatment on the high-risk patients.

Additional clinical factors can also modify the clinical benefit of remote monitoring, such as the time from implantation to initiation of remote monitoring [31], the adherence to this technique [11], or the frequency of data transmission. Two recently published meta-analyses demonstrated that significant mortality benefit was only seen in the subset of trials using a daily transmission verification (ie, Biotronik Home Monitoring System) [26,27]. A possible explanation for this difference is that the rate of events recognized within 24 hours is the highest among manufacturers with the Biotronik Home Monitoring system [32]. However, the type of remote monitoring system did not influence the survival benefit in our patient cohort.

## Limitations

This research is a single-center, retrospective study with all the consequential limitations. Despite the adjustment for the most relevant baseline cofounders, a residual selection bias can not be completely excluded. Moreover, the choice of device was not randomized but was left to the implanting physician, which may have modified our results. The patient's decision to consent to remote monitoring may also have modified the results. The potential sources of bias should also be addressed. The most important one is the patient's decision to consent with a remote monitoring program. Patients with better adherence and motivation are more likely to participate in such programs, which might have improved the outcomes of patients in the remote monitoring group.

## Conclusion

In this single-center, retrospective study of optimally treated heart failure patients undergoing CRT-D implantation, the use of remote monitoring systems was associated with significantly better survival. However, a higher number of follow-up visits in the heart failure outpatient clinic was needed, which suggests that this survival benefit could be achieved by an increased and focused effort to follow and treat high-risk patients. Our results call for further randomized studies with a standardized action plan after certain telemonitoring observations to define the optimal role of this technology in the follow-up of heart failure patients.

## Acknowledgments

This paper presents independent research. The article processing fee was funded by Biotronik.

## Conflicts of Interest

MV reports lecture and consulting fees from Boston Scientific/Minimal Invasive Technology Ltd outside the submitted work. GZD reports research grants from Boston Scientific, Biotronik, Medtronic, Lecture/Consulting Fees: Biotronik, Medtronic, St Jude Medical outside the submitted work. The other authors have no disclosures to declare.

## Multimedia Appendix 1

Definitions of clinically relevant events.

[PDF File (Adobe PDF File), 15KB - jmir_v21i7e14142_app1.pdf ]

## References

1. Lazarus A. Remote, wireless, ambulatory monitoring of implantable pacemakers, cardioverter defibrillators, and cardiac resynchronization therapy systems: analysis of a worldwide database. Pacing Clin Electrophysiol 2007 Jan;30 Suppl 1:S2-S12 [FREE Full text] [doi: 10.1111/j.1540-8159.2007.00595.x] [Medline: 17302706]
2. Varma N, Stambler B, Chun S. Detection of atrial fibrillation by implanted devices with wireless data transmission capability. Pacing Clin Electrophysiol 2005 Jan;28 Suppl 1:S133-S136 [FREE Full text] [doi: 10.1111/j.1540-8159.2005.00083.x] [Medline: 15683480]
3. Varma N, Epstein AE, Irimpen A, Schweikert R, Love C. Efficacy and safety of automatic remote monitoring for implantable cardioverter-defibrillator follow-up: the Lumos-T Safely Reduces Routine Office Device Follow-up (TRUST) trial. Circulation 2010 Jul 27;122(4):325-332 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.110.937409] [Medline: 20625110]
4. Crossley GH, Boyle A, Vitense H, Chang Y, Mead RH. The CONNECT (Clinical Evaluation of Remote Notification to Reduce Time to Clinical Decision) trial: the value of wireless remote monitoring with automatic clinician alerts. J Am Coll Cardiol 2011 Mar 8;57(10):1181-1189 [FREE Full text] [doi: 10.1016/j.jacc.2010.12.012] [Medline: 21255955]
5. Guédon-Moreau L, Lacroix D, Sadoul N, Clémenty J, Kouakam C, Hermida J, ECOST trial Investigators. A randomized study of remote follow-up of implantable cardioverter defibrillators: safety and efficacy report of the ECOST trial. Eur Heart J 2013 Feb;34(8):605-614 [FREE Full text] [doi: 10.1093/eurheartj/ehs425] [Medline: 23242192]
6. Hindricks G, Elsner C, Piorkowski C, Taborsky M, Geller JC, Schumacher B, et al. Quarterly vs yearly clinical follow-up of remotely monitored recipients of prophylactic implantable cardioverter-defibrillators: results of the REFORM trial. Eur Heart J 2014 Jan;35(2):98-105 [FREE Full text] [doi: 10.1093/eurheartj/eht207] [Medline: 23868932]
7. Landolina M, Perego GB, Lunati M, Curnis A, Guenzati G, Vicentini A, et al. Remote monitoring reduces healthcare use and improves quality of care in heart failure patients with implantable defibrillators: the evolution of management strategies of heart failure patients with implantable defibrillators (EVOLVO) study. Circulation 2012 Jun 19;125(24):2985-2992 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.088971] [Medline: 22626743]
8. Piccini JP, Mittal S, Snell J, Prillinger JB, Dalal N, Varma N. Impact of remote monitoring on clinical events and associated health care utilization: a nationwide assessment. Heart Rhythm 2016 Dec;13(12):2279-2286. [doi: 10.1016/j.hrthm.2016.08.024] [Medline: 27544748]
9. Vamos M, Nyolczas N, Bari Z, Bogyi P, Muk B, Szabo B, et al. Refined heart failure detection algorithm for improved clinical reliability of OptiVol alerts in CRT-D recipients. Cardiol J 2018;25(2):236-244 [FREE Full text] [doi: 10.5603/CJ.a2017.0077] [Medline: 28653309]
10. Saxon LA, Hayes DL, Gilliam FR, Heidenreich PA, Day J, Seth M, et al. Long-term outcome after ICD and CRT implantation and influence of remote device follow-up: the ALTITUDE survival study. Circulation 2010 Dec 07;122(23):2359-2367. [doi: 10.1161/CIRCULATIONAHA.110.960633] [Medline: 21098452]
11. Varma N, Piccini JP, Snell J, Fischer A, Dalal N, Mittal S. The relationship between level of adherence to automatic wireless remote monitoring and survival in pacemaker and defibrillator patients. J Am Coll Cardiol 2015 Jun 23;65(24):2601-2610 [FREE Full text] [doi: 10.1016/j.jacc.2015.04.033] [Medline: 25983008]

12. De Simone A, Leoni L, Luzi M, Amellone C, Stabile G, La Rocca V, et al. Remote monitoring improves outcome after ICD implantation: the clinical efficacy in the management of heart failure (EFFECT) study. Europace 2015 Aug;17(8):1267-1275. [doi: 10.1093/europace/euu318] [Medline: 25842271]
13. Hindricks G, Taborsky M, Glikson M, Heinrich U, Schumacher B, Katz A, et al. Implant-based multiparameter telemonitoring of patients with heart failure (IN-TIME): a randomised controlled trial. Lancet 2014 Aug 16;384(9943):583-590. [doi: 10.1016/S0140-6736(14)61176-4] [Medline: 25131977]
14. Boriani G, Da Costa A, Quesada A, Ricci RP, Favale S, Boscolo G, MORE-CARE Study Investigators. Effects of remote monitoring on clinical outcomes and use of healthcare resources in heart failure patients with biventricular defibrillators: results of the MORE-CARE multicentre randomized controlled trial. Eur J Heart Fail 2017 Mar;19(3):416-425. [doi: 10.1002/ejhf.626] [Medline: 27568392]
15. Morgan JM, Kitt S, Gill J, McComb JM, Ng GA, Raftery J, et al. Remote management of heart failure using implantable electronic devices. Eur Heart J 2017 Aug 07;38(30):2352-2360 [FREE Full text] [doi: 10.1093/eurheartj/ehx227] [Medline: 28575235]
16. Brignole M, Auricchio A, Baron-Esquivias G, Bordachar P, Boriani G, Breithardt O, ESC Committee for Practice Guidelines (CPG), Document Reviewers, et al. 2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: the Task Force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in collaboration with the European Heart Rhythm Association (EHRA). Eur Heart J 2013 Aug;34(29):2281-2329. [doi: 10.1093/eurheartj/eht150] [Medline: 23801822]
17. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, 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 J Heart Fail 2016 Aug;18(8):891-975 [FREE Full text] [doi: 10.1002/ejhf.592] [Medline: 27207191]
18. Crossley G, Boyle A, Vitense H, Sherfesee L, Mead RH. Trial design of the clinical evaluation of remote notification to reduce time to clinical decision: the Clinical evaluation Of remote NotificatioN to rEduCe Time to clinical decision (CONNECT) study. Am Heart J 2008 Nov;156(5):840-846. [doi: 10.1016/j.ahj.2008.06.028] [Medline: 19061696]
19. Burri H, Quesada A, Ricci RP, Boriani G, Davinelli M, Favale S, et al. The MOnitoring Resynchronization dEvices and CARdiac patiEnts (MORE-CARE) study: rationale and design. Am Heart J 2010 Jul;160(1):42-48. [doi: 10.1016/j.ahj.2010.04.005] [Medline: 20598971]
20. Guédon-Moreau L, Chevalier P, Marquié C, Kouakam C, Klug D, Lacroix D, ECOST trial Investigators. Contributions of remote monitoring to the follow-up of implantable cardioverter-defibrillator leads under advisory. Eur Heart J 2010 Sep;31(18):2246-2252 [FREE Full text] [doi: 10.1093/eurheartj/ehq203] [Medline: 20591841]
21. Arya A, Block M, Kautzner J, Lewalter T, Mörtel H, Sack S, IN-TIME investigators. Influence of Home Monitoring on the clinical status of heart failure patients: design and rationale of the IN-TIME study. Eur J Heart Fail 2008 Nov;10(11):1143-1148 [FREE Full text] [doi: 10.1016/j.ejheart.2008.08.004] [Medline: 18805053]
22. Wilkoff BL, Auricchio A, Brugada J, Cowie M, Ellenbogen KA, Gillis AM, et al. HRS/EHRA expert consensus on the monitoring of cardiovascular implantable electronic devices (CIEDs): description of techniques, indications, personnel, frequency and ethical considerations. Heart Rhythm 2008 Jun;5(6):907-925. [doi: 10.1016/j.hrthm.2008.04.013] [Medline: 18551743]
23. Slotwiner D, Varma N, Akar JG, Annas G, Beardsall M, Fogel RI, et al. HRS Expert Consensus Statement on remote interrogation and monitoring for cardiovascular implantable electronic devices. Heart Rhythm 2015 Jul;12(7):e69-e100. [doi: 10.1016/j.hrthm.2015.05.008] [Medline: 25981148]
24. Kurek A, Tajstra M, Gadula-Gacek E, Buchta P, Skrzypek M, Pyka L, et al. Impact of remote monitoring on long-term prognosis in heart failure patients in a real-world cohort: results from All-Comers COMMIT-HF trial. J Cardiovasc Electrophysiol 2017 Apr;28(4):425-431. [doi: 10.1111/jce.13174] [Medline: 28176442]
25. Klersy C, Boriani G, De Silvestri A, Mairesse GH, Braunschweig F, Scotti V, Health Economics Committee of the European Heart Rhythm Association. Effect of telemonitoring of cardiac implantable electronic devices on healthcare utilization: a meta-analysis of randomized controlled trials in patients with heart failure. Eur J Heart Fail 2016 Feb;18(2):195-204 [FREE Full text] [doi: 10.1002/ejhf.470] [Medline: 26817628]
26. Parthiban N, Esterman A, Mahajan R, Twomey DJ, Pathak RK, Lau DH, et al. Remote monitoring of implantable cardioverter-defibrillators: a systematic review and meta-analysis of clinical outcomes. J Am Coll Cardiol 2015 Jun 23;65(24):2591-2600 [FREE Full text] [doi: 10.1016/j.jacc.2015.04.029] [Medline: 25983009]
27. Hindricks G, Varma N, Kacet S, Lewalter T, Søgaard P, Guédon-Moreau L, et al. Daily remote monitoring of implantable cardioverter-defibrillators: insights from the pooled patient-level data from three randomized controlled trials (IN-TIME, ECOST, TRUST). Eur Heart J 2017 Jun 07;38(22):1749-1755 [FREE Full text] [doi: 10.1093/eurheartj/ehx015] [Medline: 29688304]
28. Yu C, Wang L, Chau E, Chan R, Kong S, Tang M, et al. Intrathoracic impedance monitoring in patients with heart failure: correlation with fluid status and feasibility of early warning preceding hospitalization. Circulation 2005 Aug 09;112(6):841-848 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.104.492207] [Medline: 16061743]

29. Böhm M, Drexler H, Oswald H, Rybak K, Bosch R, Butter C, OptiLink HF Study Investigators. Fluid status telemedicine alerts for heart failure: a randomized controlled trial. Eur Heart J 2016 Nov 01;37(41):3154-3163. [doi: 10.1093/eurheartj/ehw099] [Medline: 26984864]
30. Stockburger M, Helms TM, Perings CA, Deneke T, Köhler F, Leonhardt V, et al. Nutzenbewertung des strukturierten Telemonitorings mithilfe von aktiven Herzrhythmusimplantaten. Kardiologe 2017 Oct 23;11(6):452-459 [FREE Full text] [doi: 10.1007/s12181-017-0203-8]
31. Mittal S, Piccini J, Snell J, Prillinger J, Dalal N, Varma N. Improved survival in patients enrolled promptly into remote monitoring following cardiac implantable electronic device implantation. J Interv Card Electrophysiol 2016 Aug;46(2):129-136 [FREE Full text] [doi: 10.1007/s10840-016-0112-y] [Medline: 26860839]
32. Söth-Hansen M, Witt C, Rasmussen M, Kristensen J, Gerdes C, Nielsen J. Time until diagnosis of clinical events with different remote monitoring systems in implantable cardioverter-defibrillator patients. Heart Rhythm 2018 Dec;15(11):1648-1654. [doi: 10.1016/j.hrthm.2018.05.025] [Medline: 29803850]

## Abbreviations

**ACEi:** angiotensin-converting-enzyme inhibitor
**ARB:** angiotensin-receptor blocker
**CRT-D:** cardiac resynchronization therapy defibrillator
**HR:** hazard ratio
**ICD:** implantable cardioverter defibrillator
**NYHA:** New York Heart Association

*Edited by G Eysenbach; submitted 26.03.19; peer-reviewed by R Casado Arroyo, E van der Velde; comments to author 27.04.19; revised version received 07.05.19; accepted 07.05.19; published 26.07.19.*

*Please cite as:*
*Bogyi P, Vamos M, Bari Z, Polgar B, Muk B, Nyolczas N, Kiss RG, Duray GZ*
*Association of Remote Monitoring With Survival in Heart Failure Patients Undergoing Cardiac Resynchronization Therapy: Retrospective Observational Study*
*J Med Internet Res 2019;21(7):e14142*
*URL: http://www.jmir.org/2019/7/e14142/*
*doi:10.2196/14142*
*PMID:31350836*

Original Paper

# Effects of Digital Device Ownership on Cognitive Decline in a Middle-Aged and Elderly Population: Longitudinal Observational Study

Yinzi Jin[1], PhD; Mingxia Jing[2], PhD; Xiaochen Ma[3], PhD

[1]Department of Global Health, School of Public Health, Peking University, Beijing, China

[2]Department of Public Health, Shihezi University School of Medicine, Xinjiang, China

[3]China Center for Health Development Studies, Peking University, Beijing, China

**Corresponding Author:**
Xiaochen Ma, PhD
China Center for Health Development Studies
Peking University
38 Xueyuan Road, Haidian District
Beijing,
China
Phone: 86 10 82805706
Email: xma@hsc.pku.edu.cn

## Abstract

**Background:** Cognitive decline is a major risk factor for disability and death and may serve as a precursor of dementia. Digital devices can provide a platform of cognitively stimulating activities which might help to slow cognitive decline during the process of normal aging.

**Objective:** This longitudinal study aimed to examine the independent protective factors of desktop and cellphone ownership against cognitive decline in mid-life and older adulthood and to examine the combined effect of desktop and cellphone ownership on the same outcome.

**Methods:** Data was obtained from a China Health and Retirement Longitudinal Studies (CHARLS) cohort made up of 13,457 community-dwelling adults aged 45 years or above in 2011-2012. They were followed for 4 years, with baseline measurements taken as well as 2 two-year follow-up visits. Cognitive function was tested during the baseline test and follow-up visits. A global cognition z-score was calculated based on two domains: word recall and mental intactness. The key independent variables were defined as: whether one had desktops with internet connection at home and whether one had a cellphone. An additional categorical variable of three values was constructed as: 0 (no desktop or cellphone), 1 (desktop or cellphone alone), and 2 (desktop and cellphone both). Mixed-effects regression was adjusted for demographic and health behavior as well as health condition risk factors.

**Results:** Adjusted for demographic and health behavior as well as health condition risk factors, desktop and cellphone ownership were independently associated with subsequent decreased cognitive decline over the four-year period. Participants without a desktop at home had an adjusted cognitive decline of –0.16 standard deviations (95% CI –0.18 to –0.15), while participants with a desktop at home had an adjusted cognitive decline of –0.10 standard deviations (95% CI –0.14 to –0.07; difference of –0.06 standard deviations; $P$=.003). A similar pattern of significantly protective association of 0.06 standard deviations (95% CI 0.03-0.10; $P$<.001) between cellphone ownership and cognitive function was observed over the four-year period. Additionally, a larger longitudinal protective association on cognitive decline was observed among those with both of the digital devices, although the 95% CIs for the coefficients overlapped with those with a single digital device alone.

**Conclusions:** Findings from this study underscored the importance of digital devices as platforms for cognitively stimulating activities to delay cognitive decline. Future studies focusing on use of digital devices are warranted to investigate their longitudinal protective factors against cognitive decline at mid- and later life.

***(J Med Internet Res 2019;21(7):e14210)*** doi:10.2196/14210

**KEYWORDS**

digital access; cognitive decline; China

## Introduction

Cognitive decline is a major risk factor for physical disability [1] and death [2] and may serve as a precursor of dementia [3,4]. Cognitive decline, which often begins in individuals aged between 45-60 years, is an irreversible pathophysiological process of brain change [5]. In China, the burden of cognitive decline is increasing as a result of populations aging rapidly. Recent studies show that almost 20% of Chinese adults aged 60 or above have mild cognitive impairment, of which 6% progress to dementia annually [6-8]. However, resources for the care of those elderly with cognitive decline are insufficient, raising public health concerns about suboptimal geriatric medicine and psychiatric care services [9].

To address this challenge of cognitive decline, innovative strategies for interventions are needed. A growing number of epidemiologic studies have clearly suggested that digital devices, like internet-based instruments, can provide a platform of cognitively stimulating activities which might help to slow the path of cognitive decline in the process of normal aging [10-12]. Longitudinal studies and meta-analyses in developed countries have shown that the use of the internet can be a protective factor for cognitive decline among the older population [13-16]. In the context of China, there has been a marked increase in internet access among mid-aged and elderly populations [17], although the national data on internet use is limited. Given the potential protective association between internet use and cognitive decline, two questions remain unanswered in China. First, does the protective association between digital devices and cognitive decline found in other settings apply to mid-aged and elderly populations in China, where access to digital devices and the aging population are both growing rapidly? Second, given the diverse digital instruments that provide internet connection through which different activities might be facilitated, what is the combined effect of multiple digital instruments such as desktops and cellphones?

This study used a nationally representative longitudinal survey to: (1) investigate whether desktop and cellphone ownership might be independently associated with decreased cognitive decline in mid-life and older adulthood; and (2) to examine the combined effect of both of the digital devices on cognitive decline.

## Methods

### Design, Setting, and Participants

Participants were enrolled in the China Health and Retirement Longitudinal Study (CHARLS), a biennial prospective observational study, nationally representative of Chinese adults aged ≥45 years and their spouses. This study includes assessments of social, economic, and health circumstances of community residents from 2011-2012 [18]. Data from four years of follow-up visits (2011-2012, 2013-2014, 2015-2016) were utilized. Of the 13,548 participants who underwent cognitive assessment in the baseline wave (2011-2012), 91 participants were excluded due to lack of data on desktop and cellphone information. Among 13,457 participants enrolled as our analytic (baseline) cohort, a lot of 2073 (15.4%) were lost to follow-up due to not attending one or more follow-up surveys during the four years of the study period. Potential reasons for attrition included death or dropping out from the study. All participants provided written informed consent, and survey protocols were approved by the Peking University Ethics Review Board [19].

### Measurements

#### Cognitive Function

Cognitive functions were tested during the baseline test and during the two follow-up in-person household visits in a two-year interval. Following a previous study on CHARLS [20], the present study used a global cognitive function score based on two domains of cognitive functions: word recall and mental intactness. For word recall, each participant was asked to repeat as many of the 10 Chinese nouns just read to them as possible (immediate word recall) and then to recall the same 10-word list 5 minutes later (delayed recall). Answers to these questions were aggregated into a single word recall score ranging from 0 to 10. For the mental intactness domain, measures including numerical ability, time orientation, and picture drawing were used to formulate the score. Numerical ability was measured by serial 7 subtractions from 100 (up to five times), time orientation was measured by naming that day's date (month, day, year and date of week), and picture drawing was measured by asking each participant to redraw a picture of two overlapping pentagons shown to them. Answers to these questions were then aggregated into a single mental intactness score ranging from 0 to 10. To aid in comparison of cognitive test results, z-scores (the difference between participant's score and the sample mean, divided by the standard deviation of the baseline sample) were created for both domain-specific cognitive function of word recall and mental intactness. A global z-score was calculated according to the means and standard deviations of two domain-specific z-scores. The details of implementation procedures (interviewer recruitment, training and material preparation, quality control and data cleaning) were added in Multimedia Appendix 1.

#### Internet Access

The key independent variables were defined as whether there was the presence of desktops with internet connection at home (desktop ownership) and whether an individual had a cellphone (cellphone ownership). Because the survey did not ask whether the cellphone was a smartphone or a mobile phone without internet connection, this study's definition of cellphone ownership refers to mobile phones both with and without internet connection. Participants with both desktop and cellphone ownership, as defined above, were classified as having combined digital device ownership. Because only 1.23% (166/13,457) of our sample had desktops alone, a categorical variable of three values was then constructed as 0 (no desktop or cellphone), 1 (desktop or cellphone alone), and 2 (desktop and cellphone both).

### *Covariates*

Following previous studies, there were several covariates that were included in the present study [21-23]. They include demographic and socioeconomic status covariates such as age, gender (male or female), educational attainment (illiterate, part of primary school, primary school, middle school, high school or above), marital status (married, widowed, separated, divorced, or never married), and residence status (currently living in rural or urban area). In addition, there were health behavior covariates which included smoking status (current, former or never) and alcohol drinker (current, former or never). Lastly, there were health condition covariates included, such as whether or not the participant self-reported whether they had ever been diagnosed with the following diseases: high blood pressure, diabetes, or stroke.

### Statistical Analysis

A descriptive analysis presented the characteristics of study participants among the full sample as a whole and then by subsamples of desktop and cellphone ownership status. Two pairs of linear mixed-effect models were used to assess the longitudinal association of digital access and cognitive function over time. The first pair of models focused on desktop and cellphone ownership as independent predictors, and the second pair of models focused on the combined effect of having both desktop and cellphone or the single effect of having the desktop or cellphone alone, compared to having no desktop or cellphone. Stratified analysis by age and gender was conducted to test differential associations across subgroups. All models were adjusted for demographic (age, sex, education, marriage, rural or urban residence) and health behavior (smoke, drink) as well as health condition risk factors (self-reported hypertension, diabetes, and stroke). In order to evaluate the potential bias due to non-response and attrition across difference survey waves, sensitivity analysis was conducted using the same models with the complete sample who participated in all three waves. All analyses were conducted in Stata 14.1 (StataCorp LP).

## Results

### Characteristics of Participants

Data from 13,457 participants aged 45 or above (mean age 58.7, SD 9.37) were included for analysis. At baseline, participants with a desktop were more likely to be younger, female, better educated, either married or partnered, and less likely to be living in a rural area and smoke than participants without desktop ownership. Fewer differences were observed between participants with and without cellphones, however participants with cellphones were more likely to be younger, better educated, and either married or partnered Table 1. We also compared the baseline characteristics across three groups (without desktop or cellphone, with desktop or cellphone only, and with both desktop and cellphone) in Multimedia Appendix 2.

**Table 1.** Characteristics of the study participants.

| Characteristic | Total (N=13,457) | Desktop Ownership | | | Cellphone Ownership | | |
|---|---|---|---|---|---|---|---|
| | | No (n=11,143) | Yes (n=2314) | *P* value | No (n=2764) | Yes (n=10,693) | *P* value |
| Age, mean (SD) | 58.7 (9.37) | 59.3 (9.43) | 55.6 (8.44) | <.001 | 64.8 (9.42) | 57.1 (8.68) | <.001 |
| Male sex, n (%) | 6590 (49.0) | 5508 (49.4) | 1082 (46.8) | .02 | 1339 (48.4) | 5251 (49.1) | .53 |
| **Educational level, n (%)** | | | | **<.001** | | | **<.001** |
| Illiterate | 3340 (24.8) | 3106 (27.9) | 234 (10.1) | | 1133 (41.0) | 2207 (20.6) | |
| Part of primary school | 2462 (18.3) | 2208 (19.8) | 254 (11.0) | | 562 (20.3) | 1900 (17.8) | |
| Primary school | 2937 (21.8) | 2549 (22.9) | 388 (16.8) | | 543 (19.7) | 2394 (22.4) | |
| Middle school | 2924 (21.7) | 2251 (20.2) | 673 (29.1) | | 344 (12.5) | 2580 (24.1) | |
| High school or above | 1793 (13.3) | 1029 (9.2) | 764 (33.0) | | 181 (6.6) | 1612 (15.1) | |
| **Marital status, n (%)** | | | | **<.001** | | | **<.001** |
| Married or Partnered | 11,863 (88.2) | 9717 (87.2) | 2146 (92.7) | | 2227 (80.6) | 9636 (90.1) | |
| Otherwise | 1594 (11.8) | 1426 (12.8) | 168 (7.3) | | 537 (19.4) | 1057 (9.9) | |
| Rural residence, mean (SD) | 0.59 (0.49) | 0.66 (0.47) | 0.26 (0.44) | <.001 | 0.66 (0.48) | 0.57 (0.49) | <.001 |
| **Smoke, n (%)** | | | | **<.001** | | | **.69** |
| Current | 4205 (31.3) | 3578 (32.1) | 627 (27.1) | | 844 (30.5) | 3361 (31.4) | |
| Former | 1237 (9.2) | 1035 (9.3) | 202 (8.7) | | 276 (10.0) | 961 (9.0) | |
| Never | 8013 (59.6) | 6528 (58.6) | 1485 (64.2) | | 1643 (59.5) | 6370 (59.6) | |
| **Alcohol drinker, n (%)** | | | | **<.001** | | | **.01** |
| Current | 4492 (33.4) | 3684 (33.1) | 808 (34.9) | | 856 (31.0) | 3636 (34.0) | |
| Former | 1140 (8.5) | 987 (8.9) | 153 (6.6) | | 262 (9.5) | 878 (8.2) | |
| Never | 7821 (58.1) | 6468 (58.1) | 1353 (58.5) | | 1646 (59.6) | 6175 (57.8) | |
| Ever had high blood pressure, mean (SD) | 0.3 (0.44) | 0.3 (0.44) | 0.3 (0.44) | .87 | 0.3 (0.46) | 0.3 (0.44) | <.001 |
| Ever had diabetes, mean (SD) | 0.1 (0.25) | 0.1 (0.24) | 0.1 (0.27) | .01 | 0.1 (0.25) | 0.1 (0.25) | .33 |
| Ever had stroke, mean (SD) | 0.0 (0.16) | 0.0 (0.16) | 0.0 (0.15) | .26 | 0.0 (0.17) | 0.0 (0.15) | .02 |

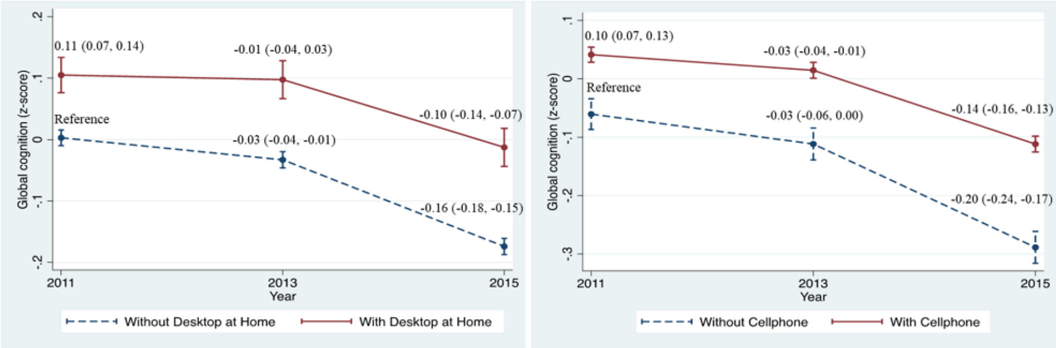
### Independent Association Between Desktop or Cellphone Ownership and Cognitive Function Over Time

In mixed-effects regression adjusted for demographic and health behavior as well as health condition risk factors, desktop or cellphone ownership were associated with similarly higher baseline global cognitive scores. Compared with participants without a desktop at home, participants with desktop ownership had global cognitive scores at baseline that were 0.11 standard deviations (95% CI 0.07-0.14) higher. Similar positive association was found between cellphone ownership and baseline global cognitive scores of 0.10 (95% CI 0.07-0.13) (Figure 1).

We then investigated whether desktop or cellphone ownership was independently associated with subsequent cognitive trajectories, adjusted for demographic and health behavior as well as health condition risk factors. Compared to participants without desktop ownership, there was a significant reduction in cognitive decline among participants with desktop ownership in the four-year follow-up. On average, over the two-year follow-up, participants without a desktop at home had an adjusted cognitive decline of –0.03 standard deviations (95% CI –0.04 to –0.01), while participants with a desktop at home had an adjusted cognitive decline of –0.01 standard deviations (95% CI –0.04 to 0.03). However, this difference of –0.02 standard deviations was not significant ($P$=.12). Over the four-year follow-up, participants without a desktop at home had an adjusted cognitive decline of –0.16 standard deviations (95% CI –0.18 to –0.15), while participants with a desktop at home had an adjusted cognitive decline of –0.10 standard deviations (95% CI –0.14 to –0.07; difference of –0.06 standard deviations; $P$=.003) (Figure 1).

A similar pattern of significantly protective association of 0.06 standard deviations (95% CI 0.03-0.10; $P$<.001) between cellphone ownership and cognitive function was found over the four-year period while no significant difference was observed in the two-year follow-up period (Figure 1).

**Figure 1.** Independent association between desktop or cellphone ownership and cognitive function over time. Adjusted for demographic (age, sex, education, marriage, rural or urban residence) and health behavior (smoke, drink) as well as health condition risk factors (self-reported hypertension, diabetes, and stroke).




Panel A. Desktop Ownership and Cognitive Decline

Panel B. Cellphone Ownership and Cognitive Decline

### Combined Effect of Desktop and Cellphone Ownership on Cognitive Function Over Time

In the next subsection we investigated the combined effect of desktop and cellphone ownership, adjusted for demographic and health behavior as well as health condition risk factors. In the baseline measured global cognitive score, participants having desktop or cellphone alone were 0.10 standard deviations (95% CI 0.07-0.13) higher and participants having both desktop and cellphone were 0.20 standard deviations (95% CI 0.15-0.24) higher than participants having neither desktop nor cellphone (Table 2).

Similar to the patterns in the independent association between desktop or cellphone ownership and cognitive decline, the longitudinal protective association was found in the longer-term four-year follow-up. Compared to participants having neither desktop nor cellphone, greater longitudinal protective association was observed among participants having both desktop and cellphone (SD 0.10, 95% CI 0.05-0.15). Although the effect size was larger than those participants having desktop or cellphone alone (SD 0.07, 95% CI 0.03-0.10), the two 95% CIs overlapped with each other (Table 2).

**Table 2.** Adjusted longitudinal association between digital device ownership and cognitive function over time (Total Number of Observations=33,956).

| Variables | Point Estimates (95% CI) | *P* value |
|---|---|---|
| **Baseline Cognitive Function** | | |
| No Desktop or Cellphone | reference | |
| Desktop or Cellphone Alone | 0.10 (0.07-0.13) | <.001 |
| Desktop and Cellphone Both | 0.20 (0.15-0.24) | <.001 |
| **Longitudinal Protective Association in 2 Years** | | |
| No Desktop or Cellphone | reference | |
| Desktop or Cellphone Alone | 0.02 (–0.02 to 0.05) | .39 |
| Desktop and Cellphone Both | 0.02 (–0.02 to 0.08) | .30 |
| **Longitudinal Protective Association in 4 Years** | | |
| No Desktop or Cellphone | reference | |
| Desktop or Cellphone Alone | 0.07 (0.03-0.10) | <.001 |
| Desktop and Cellphone Both | 0.10 (0.05-0.15) | <.001 |

### Stratified Analysis and Sensitivity Analysis

Our stratified analysis found that the protective association between digital device ownership and cognitive decline during the four-year follow-up was primarily driven by females (Multimedia Appendix 3) and relatively younger adult participants aged 45-59 (Multimedia Appendix 4). Sensitivity analyses restricted analytic cohorts to those 11,384 who participated in all survey waves. We found similar results with the original analysis (Multimedia Appendix 5).

## *Discussion*

### Principal Findings

This study is, to our knowledge, the first nationally representative study to show the longitudinal protective association between digital device ownership and cognitive function among mid-aged and elderly populations in China. China has the largest population aged 65 years and older in the world, and in 2050 the proportion of older persons within the total national population is projected to be around 25% [24]. The health needs for elderly people with irreversible cognitive impairment challenge China's health and social care system in a serious and unprecedented way. Over the past decade the proportion of the population connected to the internet has been growing exponentially, and digital device ownership shows potential for older people because it enables easier access to experiential, intelligential, social or emotional support for health purposes [25]. Taking advantage of the population-based nationally representative survey, we were able to investigate whether the preventive effect of desktop or cellphone ownership on cognitive function found in other settings still holds in the context of China, where access to internet and population aging are simultaneously increasing, and whether different digital devices have a combined effect on cognitive decline. We believed our results were less likely to be subject to non-response or attrition bias because our sensitivity analysis yielded similar results when using whole sample and complete cases samples.

We found that individuals with digital device ownership had better cognitive function at baseline and a lower rate of decline than those without internet devices in a four-year follow-up. Our finding is consistent with a study using the English Longitudinal Study of Aging (ELSA) cohort which showed that internet users had better cognitive performance measured by delayed recall compared with internet nonusers [13]. Other results of a protective association between internet use and cognitive function were also reported in other developed countries [12-16]. Although the exact mechanism is not clear, a potential theoretical pathway may be that internet access imparted a reserve against the expression of cognitive impairment [10,26,27]. The cognitively stimulating activities facilitated by digital devices generate a cognitive reserve buffer that promotes resilience so as to prevent or delay the development of cognitive impairment [28]. In addition, we found that the protective effect was larger during the four-year follow-up than that during the two-year follow-up. This is in line with previous studies that show that longer and more intense interventions aimed at the use of digital devices might significantly enhance cognitive functions [29-31].

Although not statistically significant, we observed a trend towards greater longitudinal protective association between cognitive function and both of the digital instruments relative to access to a single device (desktop or cellphone ownership) alone. Different digital devices have different functions and areas of cognition-related application. Computers are mainly used to search for health-related information, such as consulting a doctor at a distance, thus enabling easier access to better and more effective health care for adults [32]. In the context of China, older people tend to use the internet for obtaining health information about proper nutrition, exercise or weight issues and disease management, which contributes to adopting healthier behaviors and making more informed medical decisions. However, cellphones are mainly used to stay in contact with friends, engage in entertainment activities, receive reminders for scheduled visits and for medication instructions, which therefore increase social participation and interaction. Meanwhile, digital device ownership in general stimulates the learning of new things and increases the cognitive demand to master new skills using different digital tools. Future studies focusing on multiple digital devices are warranted to assess the longitudinal protective association between digital devices and cognitive function at mid- and later life.

Stratified analyses indicated that gender and age modified protective association between digital device ownership and cognitive decline, which is consistent with previous studies in other settings [33-36]. Our results suggest that targeting policies on cognitive impairment and dementia prevention around females and earlier age groups might improve their effectiveness. Future studies should focus on understanding mechanisms through which digital device ownership works on slowing cognitive decline.

In summary, our encouraging findings highlighted the importance of promoting the application of internet-based computers, cellphones and other digital devices in middle-aged and elderly Chinese. As more people are connected with multiple digital instruments, the benefit of digital device ownership can be maximized for clinical and health services delivery. A growing group of studies have demonstrated the effectiveness of mobile-based health interventions, such as online cognitive training programs, internet-based conversations and internet or email use [29,30]. A majority of older Chinese have owned a mobile phone at some point, and this high ownership rate suggests that it could become a tool that can be implemented into clinical interventions to reduce cognition loss, especially when connected with the internet. More broadly, our results might promote more widely accessible mobile phones and the continuing penetration of internet access, in order to tackle the increasing challenges of aging in China.

### Limitations

Our study had several limitations. First, the observational nature of our study limited our ability to investigate the causal relationship between digital device ownership and cognitive decline. The results should not be interpreted as the long-term effect of digital device ownership on reducing cognitive decline. Our results also didn’t determine any mechanistic basis behind the observed link between digital device ownership and subsequent cognitive trajectories. Rather, the longitudinal protective association between digital device ownership and cognitive decline found in the present study underscored the need for research to capitalize on new digital technologies to slow cognitive decline.

Second, due to data feasibility, our analysis was limited to digital device ownership without taking into account digital device use. This might limit our ability to understand how digital devices work against cognitive decline. Nevertheless, previous studies have shown that digital device ownership demonstrated a high correlation with use of digital devices [37,38].

Third, because of data availability, the definition of cellphone ownership in our study referred to a mobile phone both with and without internet connection. This might limit our ability to explore the mechanism via which (internet-based or not) the cellphone had a protective association with cognition. Meanwhile, we highlighted that different digital devices have different functions and areas of cognition-related application. Computers are mainly used to search online for health-related information, while cellphones are mainly used to increase social participation and interaction via telephone calls and messages. Interestingly, we found that the magnitude of protective association between computers and cognition was about the same as that between cellphones and cognition.

Fourth, our digital device ownership was only measured at baseline so information on the potential change of digital device ownership was not available. Nevertheless, the fact that our data already had a high baseline ownership rate of single devices (65%) and of both of the devices (16%), that made it less likely that the limitation of only measuring digital access at baseline would lead to differential bias in our results.

## Conclusions

Previous studies using CHARLS cohorts, including one of our own, have demonstrated digital access is associated with better physical health and better outcomes of chronic disease management [39-41]. The national representativeness of CHARLS adds to the robustness of the results, indicating digital devices as a platform for health management. This present study found the longitudinal protective association between digital device ownership and cognitive function. Findings from this study underscored the importance of digital devices as a platform for cognitively stimulating activities to delay cognitive decline. Future studies focusing on use of digital devices are warranted so as to investigate digital devices as a protective factor against cognitive decline at mid- and later life.

## Acknowledgments

The authors would like to acknowledge the CHARLS team for providing data and the training necessary to use the dataset. This study was supported by the Peking University Health Science Center start-up fund (grant no. BMU20160514).

## Authors' Contributions

YJ analyzed the data and drafted the manuscript. MJ revised the manuscript. XM designed the study, analyzed the data, and revised the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Details of interviewer recruitment, training and material preparation, quality control and data cleaning of CHARLS.

[PDF File (Adobe PDF File), 77KB - jmir_v21i7e14210_app1.pdf ]

## Multimedia Appendix 2

Characteristics of the three groups of participants by digital device ownership.

[PDF File (Adobe PDF File), 50KB - jmir_v21i7e14210_app2.pdf ]

## Multimedia Appendix 3

Adjusted association between digital device ownership and cognitive function by gender.

[PDF File (Adobe PDF File), 78KB - jmir_v21i7e14210_app3.pdf ]

## Multimedia Appendix 4

Adjusted association between digital device ownership and cognitive function by age.

[PDF File (Adobe PDF File), 73KB - jmir_v21i7e14210_app4.pdf ]

## Multimedia Appendix 5

Sensitivity analyses restricted analytic cohorts to those participated in all survey waves.

[PDF File (Adobe PDF File), 58KB - jmir_v21i7e14210_app5.pdf ]

## References

1. Prince MJ, Wu F, Guo Y, Gutierrez Robledo LM, O'Donnell M, Sullivan R, et al. The burden of disease in older people and implications for health policy and practice. The Lancet 2015 Feb;385(9967):549-562. [doi: 10.1016/S0140-6736(14)61347-7]

2. Huynh QL, Negishi K, Blizzard L, Saito M, De Pasquale CG, Hare JL, et al. Mild cognitive impairment predicts death and readmission within 30 days of discharge for heart failure. International Journal of Cardiology 2016 Oct;221:212-217. [doi: 10.1016/j.ijcard.2016.07.074] [Medline: 27404677]
3. Bradburn S, Murgatroyd C, Ray N. Neuroinflammation in mild cognitive impairment and Alzheimer's disease: A meta-analysis. Ageing Research Reviews 2019 Mar;50:1-8. [doi: 10.1016/j.arr.2019.01.002] [Medline: 30610927]
4. Shah Y, Tangalos E, Petersen R. Mild cognitive impairment. When is it a precursor to Alzheimer's disease? Geriatrics 2000;55:62-65. [Medline: 10997127]
5. Singh-Manoux A, Kivimaki M, Glymour MM, Elbaz A, Berr C, Ebmeier KP, et al. Timing of onset of cognitive decline: results from Whitehall II prospective cohort study. BMJ 2012 Jan 05;344:d7622 [FREE Full text] [doi: 10.1136/bmj.d7622] [Medline: 22223828]
6. Ding D, Zhao Q, Guo Q, Meng H, Wang B, Luo J, et al. Prevalence of mild cognitive impairment in an urban community in China: a cross-sectional analysis of the Shanghai Aging Study. Alzheimers Dement 2015 Mar;11(3):300-309. [doi: 10.1016/j.jalz.2013.11.002] [Medline: 24613707]
7. Ding D, Zhao Q, Guo Q, Liang X, Luo J, Yu L, Shanghai Aging Study (SAS). Progression and predictors of mild cognitive impairment in Chinese elderly: A prospective follow-up in the Shanghai Aging Study. Alzheimers Dement (Amst) 2016;4:28-36 [FREE Full text] [doi: 10.1016/j.dadm.2016.03.004] [Medline: 27489876]
8. Zeng Y, Feng Q, Hesketh T, Christensen K, Vaupel JW. Survival, disabilities in activities of daily living, and physical and cognitive functioning among the oldest-old in China: a cohort study. The Lancet 2017 Apr;389(10079):1619-1629. [doi: 10.1016/S0140-6736(17)30548-2]
9. Wang X, Chen P. Population ageing challenges health care in China. The Lancet 2014 Mar;383(9920):870-871. [doi: 10.1016/S0140-6736(14)60443-8]
10. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. The Lancet Neurology 2012 Nov;11(11):1006-1012. [doi: 10.1016/S1474-4422(12)70191-6]
11. Plassman BL, Williams JWJ, Burke JR, Holsinger T, Benjamin S. Systematic Review: Factors Associated With Risk for and Possible Prevention of Cognitive Decline in Later Life. Annals of internal medicine 2010 Aug 03;153(3):182-193. [doi: 10.7326/0003-4819-153-3-201008030-00258]
12. Klimova B. Use of the Internet as a prevention tool against cognitive decline in normal aging. Clinical interventions in aging 2016 Sep;11:1231-1237. [doi: 10.2147/CIA.S113758]
13. Xavier AJ, d'Orsi E, de Oliveira CM, Orrell M, Demakakos P, Biddulph JP, et al. English Longitudinal Study of Aging: Can Internet/E-mail Use Reduce Cognitive Decline? The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 2014 Aug;69(9):1117-1121. [doi: 10.1093/gerona/glu105]
14. Almeida OP, Yeap BB, Alfonso H, Hankey GJ, Flicker L, Norman PE. Older Men Who Use Computers Have Lower Risk of Dementia. PLoS ONE 2012 Aug;7(8):e44239. [doi: 10.1371/journal.pone.0044239]
15. Tun PA, Lachman ME. The association between computer use and cognition across adulthood: Use it so you won't lose it? Psychology and Aging 2010 Sep;25(3):560-568. [doi: 10.1037/a0019543]
16. Liapis J, Harding KE. Meaningful use of computers has a potential therapeutic and preventative role in dementia care: A systematic review. Australasian journal on ageing 2017 Jun;36(4):299-307. [doi: 10.1111/ajag.12446]
17. Liang P, Guo S. Social interaction, Internet access and stock market participation—An empirical study in China. Journal of Comparative Economics 2015 Nov;43(4):883-901. [doi: 10.1016/j.jce.2015.02.003]
18. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS). International Journal of Epidemiology 2012 Dec;43(1):61-68. [doi: 10.1093/ije/dys203]
19. Zhao Y, Strauss J, Yang G, Giles J, Hu P, Hu Y. China Health and Retirement Longitudinal Study—2011-2012 National Baseline Users' Guide. Beijing: Peking University; Feb 2013.
20. Wang T, Wu Y, Sun Y, Zhai L, Zhang D. A Prospective Study on the Association between Uric Acid and Cognitive Function among Middle-Aged and Older Chinese. Journal of Alzheimers Disease 2017 May;58(1):79-86. [doi: 10.3233/JAD-161243]
21. Poels MM, van Oijen M, Mattace-Raso FU, Hofman A, Koudstaal PJ, Witteman JC, et al. Arterial stiffness, cognitive decline, and risk of dementia: the Rotterdam study. Stroke 2007 Mar;38(3):888-892. [doi: 10.1161/01.STR.0000257998.33768.87]
22. Brodaty H, Farrow M, Flicker L, Hecker J, Velandai S. Dementia risk reduction: the evidence. Dementia 2007;13:1-20.
23. Wendell C, Gunstad J, Waldstein S, Wright J, Ferrucci L, Zonderman A. Cardiorespiratory fitness and accelerated cognitive decline with aging. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences 2014;69:455-462. [doi: 10.1093/gerona/glt144]
24. The Lancet. Ageing in China: a ticking bomb. The Lancet 2016 Oct;388(10056):2058. [doi: 10.1016/S0140-6736(16)32058-X]
25. Siliquini R, Ceruti M, Lovato E, Bert F, Bruno S, De Vito E, et al. Surfing the internet for health information: an italian survey on use and population choices. BMC medical informatics and decision making 2011 Apr;11(1):21-30. [doi: 10.1186/1472-6947-11-21]
26. Mather M, Harley CW. The Locus Coeruleus: Essential for Maintaining Cognitive Function and the Aging Brain. Trends in Cognitive Sciences 2016 Mar;20(3):214-226. [doi: 10.1016/j.tics.2016.01.001]

27. Kim GH, Jeon S, Im K, Kwon H, Lee BH, Kim GY, et al. Structural Brain Changes after Traditional and Robot-Assisted Multi-Domain Cognitive Training in Community-Dwelling Healthy Elderly. PLoS ONE 2015 Apr;10(4):e0123251. [doi: 10.1371/journal.pone.0123251]
28. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. The Lancet 2017 Dec;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6]
29. Corbett A, Owen A, Hampshire A, Grahn J, Stenton R, Dajani S, et al. The Effect of an Online Cognitive Training Package in Healthy Older Adults: An Online Randomized Controlled Trial. Journal of the American Medical Directors Association 2015 Nov;16(11):990-997. [doi: 10.1016/j.jamda.2015.06.014] [Medline: 26543007]
30. Bozoki A, Radovanovic M, Winn B, Heeter C, Anthony JC. Effects of a computer-based cognitive exercise program on age-related cognitive decline. Archives of Gerontology and Geriatrics 2013 Jul;57(1):1-7. [doi: 10.1016/j.archger.2013.02.009]
31. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS medicine 2014 Nov;11(11):e1001756. [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
32. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. Journal of medical Internet research 2015 Jan;17(1):e10. [doi: 10.2196/jmir.3749] [Medline: 25574815]
33. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. Journal of medical Internet research 2009 Feb;11(1):e4. [doi: 10.2196/jmir.1035] [Medline: 19275980]
34. Xavier AJ, d'Orsi E, Wardle J, Demakakos P, Smith SG, von Wagner C. Internet use and cancer-preventive behaviors in older adults: findings from a longitudinal cohort study. Cancer Epidemiology, Biomarkers & Prevention 2013 Nov;22(11):2066-2074 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-0542] [Medline: 24148971]
35. Kobayashi LC, Wardle J, von Wagner C. Internet use, social engagement and health literacy decline during ageing in a longitudinal cohort of older English adults. Journal of Epidemiology and Community Health 2015 Mar;69(3):278-283 [FREE Full text] [doi: 10.1136/jech-2014-204733] [Medline: 25428933]
36. Batty D, Blake M, Bridges S, Crawford R, Demakakos P, de Oliviera C, et al. In: Banks J, Nazroo J, Steptoe A, editors. The Dynamics of Ageing: Evidence from the English Longitudinal Study of Ageing 2002-2012 (Wave 6). London: Institute of Fiscal Studies; Oct 2014.
37. Kalichman SC, Weinhardt L, Benotsch E, DiFonzo K, Luke W, Austin J. Internet access and Internet use for health information among people living with HIV–AIDS. Patient Education and Counseling 2002 Feb;46(2):109-116. [doi: 10.1016/S0738-3991(01)00134-3]
38. Kenski K, Stroud NJ. Connections Between Internet Use and Political Efficacy, Knowledge, and Participation. Journal of broadcasting & electronic media 2006 Jun;50(2):173-192. [doi: 10.1207/s15506878jobem5002_1]
39. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes, Obesity and Metabolism 2009 Oct;11(10):913-930. [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]
40. Jaana MPG, Paré G, Sicotte C. Home telemonitoring for respiratory conditions: a systematic review. American journal of managed care 2009 May;15(5):313-320. [doi: 10.1016/j.ijmedinf.2006.09.001] [Medline: 19435399]
41. Jin Y, Jing M, Zhang L, Song S, Ma X. Internet Access and Hypertension Management Among the Elderly Population: A Nationally Representative Cross-Sectional Survey in China. Journal of medical Internet research 2019 Jan;21(1):e11280. [doi: 10.2196/11280]


## Abbreviations

**CHARLS:** China Health and Retirement Longitudinal Study
**ELSA:** English longitudinal Study of Aging


*Edited by G Eysenbach; submitted 30.03.19; peer-reviewed by F Lanfranchi, C Li; comments to author 14.05.19; revised version received 02.06.19; accepted 18.06.19; published 29.07.19.*

*Please cite as:*
*Jin Y, Jing M, Ma X*
*Effects of Digital Device Ownership on Cognitive Decline in a Middle-Aged and Elderly Population: Longitudinal Observational Study*
*J Med Internet Res 2019;21(7):e14210*
*URL: http://www.jmir.org/2019/7/e14210/*
*doi:10.2196/14210*
*PMID:31359864*

Original Paper

# Clinician Job Searches in the Internet Era: Internet-Based Study

Shalu Gillum[1*], JD, MLS, AHIP; Natasha Williams[1*], MLIS, AHIP; Brittany Brink[1*], MSM, PHR, SHRM-CP; Edward Ross[1*], MD

College of Medicine, University of Central Florida, Orlando, FL, United States

[*]all authors contributed equally

**Corresponding Author:**
Edward Ross, MD
College of Medicine
University of Central Florida
6850 Lake Nona Blvd
Orlando, FL,
United States
Phone: 1 407 266 1060
Email: edward.ross@ucf.edu

## Abstract

**Background:** Traditional methods using print media and commercial firms for clinician recruiting are often limited by cost, slow pace, and suboptimal results. An efficient and fiscally sound approach is needed for searching online to recruit clinicians.

**Objective:** The aim of the study was to assess the Web-based methods by which clinicians might be searching for jobs in a broad range of specialties and how academic medical centers can advertise clinical job openings to prominently appear on internet searches that would yield the greatest return on investment.

**Methods:** We used a search engine (Google) to identify 8 query terms for each of the specialties and specialists (eg, dermatology and dermatologist) to determine internet job search methodologies for 12 clinical disciplines. Searches were conducted, and the data used for analysis were the first 20 results.

**Results:** In total, 176 searches were conducted at varying times over the course of several months, and 3520 results were recorded. The following 4 types of websites appeared in the top 10 search results across all specialties searched, accounting for 52.27% (920/1760) of the results: (1) a single no-cost job aggregator (229/1760, 13.01%); (2) 2 prominent journal-based paid digital job listing services (157/1760, 8.92% and 91/1760, 5.17%, respectively); (3) a fee-based Web-based agency (137/1760, 7.78%) offering candidate profiles; and (4) society-based paid advertisements (totaling 306/1760, 17.38%). These sites accounted for 75.45% (664/880) of results limited to the top 5 results. Repetitive short-term testing yielded similar results with minor changes in the rank order.

**Conclusions:** On the basis of our findings, we offer a specific financially prudent internet strategy for both clinicians searching the internet for employment and employers hiring clinicians in academic medical centers.

*(J Med Internet Res 2019;21(7):e12638)* doi:10.2196/12638

**KEYWORDS**

personnel selection; internet; academic medical centers

## Introduction

### Background

The advent of the internet's rapid widespread communication capabilities has dramatically altered job recruitment and search methodologies from the perspectives of both the potential employee and employer. As an academic medical center with many open positions across multiple disciplines, we were challenged to update our recruiting approach to reach wide audiences in an efficient and fiscally prudent manner. In a Web-based era that permits free or low-cost postings, we questioned the value of using print advertisements, postal mailings, and expensive contracted subscription or search firm recruiting services. With maturation of Web-based job recruitment and marketing, it behooves employers from multiple industries to not only invest resources in looking for prospective employees on the internet but also to develop a strategy to assure that their websites appear prominently in Web-based searches. As only a few search engines (eg, Google) are the starting point for queries by most internet users, regardless of what is being searched [1], it is important to ascertain the method of job posting most likely to yield a display on the first couple of Web pages.

### Objectives

Our College of Medicine's recent experience was that traditional approaches of using print media and commercial recruiting firms were not only limited by budget constraints but also failed to yield optimal and rapid results. Thus, we looked at the literature seeking guidance as to the best approach for specialty and subspecialty clinician searches using the internet. In that no publications were identified that directly addressed this question, we sought to rigorously study Web-based recruitment search techniques from the viewpoint of both the employer and job seeker. We hypothesized that robust internet search engines would be able to identify job opportunities posted at no cost by employers; however, it was not clear how prominently those positions would be displayed when compared with those advertised and promoted by fee-based listing agencies. We assembled a team of faculty medical librarians, a human resource professional, and a physician to assess Web-based methods by which clinicians might be looking for jobs in a broad range of specialties and how the college should advertise clinical job openings that would yield the greatest return on investment.

## Methods

### Identifying Search Terms

To determine how clinicians could look for jobs using the internet, we selected Google for the searches based on its ranking as the most popular engine with approximately 70% of the market share [1]. To assure that we were selecting search terms that would yield results relevant to legitimate medical employment, we explored the wording and phrases that an individual might use for a query in just 2 of the clinical specialties (dermatology and endocrinology). We identified 8 terms: *jobs*, *positions*, *job openings*, *job postings*, *job advertisements*, *local Orlando jobs*, *job listings*, and *employment*. The designation for the medical field was inserted before each of these search terms, using both the name of the specialty and name of the specialist, for example, *endocrinology* or *endocrinologist*. This yielded a total of 16 searches per specialty (Table 1). As the search terms did not involve personal identifiers and the data were all in the public domain, the project is not considered human research and thus not processed by our institutional review board.

### Data Collection and Analysis

Searches were conducted, and the results were recorded from the first 2 pages of Google screens based on research that users are unlikely to view search engine results past the second page [2]. We excluded paid advertisements and image results for our analyses. Searches were conducted during weekdays at varying dates and times between the hours of 10 am and 5 pm from November 2014 to March 2015.

After the query language was finalized, the searches were then expanded to a total of 12 disciplines by including the 10 additional clinical specialties of rheumatology, podiatry, internal medicine, gastroenterology, ophthalmology, infectious disease, allergy, nephrology, neurology, and pulmonary. Searches for internal medicine and infectious disease, however, had only 8 search results each (rather than 16), as there was no appropriate *-ology* or *-ologist* suffix for those specialties (Table 1). Otherwise, the same methodology and terms were used for all these searches, and the top 20 results from each of them were recorded from the first 2 pages of results displayed by Google. We then categorized and pursued all those links to determine what the job seeker would discover at the respective websites. The links to job positions were classified as being to (1) job aggregators pulling opportunities from multiple Web sources that were of no cost to either the employer or the potential employee (eg, NoCoAg); (2) fee- or subscription-based agencies charging the employer for inclusion in their database and sometimes inviting job seekers to register their interests (eg, SubscrAg); (3) journal-based paid digital job listings (eg, SubscrJ), distinct from the option of also having them in the paper publication; and (4) society-based paid advertisements (eg, SubscrSoc). We then determined patterns to these search findings across all 12 specialties, especially as to whether paying for a listing resulted in a more prominent Web page display (eg, first screen or the top 5, 10, or 20 *hits*) compared with those from no-cost aggregators.

**Table 1.** Examples of keyword specialty search terms.

| Specialty search: endocrinology | Specialty search: infectious diseases |
|---|---|
| Jobs AND endocrinology OR endocrinologist | Jobs AND infectious disease |
| Positions AND endocrinology OR endocrinologist | Positions AND infectious disease |
| Job openings AND endocrinology OR endocrinologist | Job openings AND infectious disease |
| Job postings AND endocrinology OR endocrinologist | Job postings AND infectious disease |
| Job advertisements AND endocrinology OR endocrinologist | Job advertisements AND infectious disease |
| Local Orlando jobs AND endocrinology OR endocrinologist | Local Orlando jobs AND infectious disease |
| Job listings AND endocrinology OR endocrinologist | Job listings AND infectious disease |
| Employment AND endocrinology OR endocrinologist | Employment AND infectious disease |

### Reproducibility

There was a concern that because of the fluidity of data on the internet and unknown factors driving proprietary search engine logic, there might be considerable differences in search results over the short term. To explore this possibility, the phrase *endocrinology jobs* was randomly chosen as a test search to be repeated. These keywords were entered into Google 100 times on a Monday and another 100 times on a Tuesday during 1 week in March 2015, for a total of 200 searches. The top 20 results (ie, encompassing the first 2 search results page in Google) were recorded for every repeated query.

## Results

### Overview

In total, 176 searches were conducted for the 12 disciplines using the 8 keyword terms. The results for the first 2 pages of every search were recorded: as Google defaults to 10 results per page, this represented the top 20 for each specialty and yielded 3520 results.

There were 1760 total search results for all search terms across all specialties in the top 10 results in Google (see Multimedia Appendix 1).

### Patterns Across Search Results

Unexpectedly, our findings revealed that the following 4 websites or types of websites appeared in this first page of search results across all specialties searched: a NoCoAg, a SubscrAg, at least one SubscrJ, and a SubscrSoc website (ie, relevant to the specialty searched, such as the American Association of Clinical Endocrinologists). Of these results, a particular NoCoAg (Indeed) was identified 229 times in the top 10 (13.01% (229/1760) of the time), a SubscrAg (Practice Link) 7.78% (137/1760), 2 SubscrJ sites 8.92% (157/1760) and 5.17% (91/1760; *JAMA* Career Center and *NEJM* Career Center, respectively), and, finally, various society websites each typically in single digits but totaling 17.38% (306/1760). Combined, these websites comprise 52.27% (920/1760) of the top 10 search results for all terms searched in Google (Table 2). For-hire search firms rarely appeared in any of these queries.

The pattern was even more pronounced for the top 5 search results. There were 880 total search results in the top 5 for all search terms and across all specialties. The NoCoAg appeared 14.8% (130/880) of the time in the top 5, the SubscrAg 13.5% (119/880) the 2 SubscrJs 14.2% (125/880) and 7.6% (67/880), and the society websites totaled 25.3% (223/880). Combined, these comprise 75.4% (664/880) of the top 5 search results for all terms searched in Google (Table 3 and Multimedia Appendix 2).

**Table 2.** Top 10 search results of all specialties searched (N=1760).

| Search result (type) | Search result (website) | Appearing in top 10 search results, n (%) |
|---|---|---|
| Society websites | Multiple, combined | 306 (17.38) |
| No cost aggregator | Indeed | 229 (13.01) |
| Subscription journal job e-listing | *JAMA* Career Center | 157 (8.92) |
| Subscription aggregator | Practice Link | 137 (7.78) |
| Subscription journal job e-listing | *NEJM* Career Center | 91 (5.17) |

**Table 3.** Top 5 search results of all specialties searched (N=880).

| Search result (type) | Search (website) result | Appearing in top 5 search results, n (%) |
|---|---|---|
| Society websites | Multiple, combined | 223 (25.3) |
| No cost aggregator | Indeed | 130 (14.8) |
| Subscription journal job e-listing | *JAMA* Career Center | 125 (14.2) |
| Subscription aggregator | Practice Link | 119 (13.5) |
| Subscription journal job e-listing | *NEJM* Career Center | 67 (7.6) |

### Reproducibility

When the term *endocrinology jobs* was searched repeatedly, results on either a Monday or Tuesday did not vary. The same pattern as described above was evident. Google results were identical for each day's 100 searches. The order of the search results varied slightly but not enough to break the pattern as described.

## Discussion

### Overview

Academic centers, especially the state-regulated institutions, often face recruitment challenges not encountered in the private or retail sectors. Similar to most others, our college's human resources policies and procedures require that all faculty positions, including clinical ones, have a nationwide search and be advertised for a proscribed length of time. It is important that the process reaches a wide and diverse enough audience to meet equal opportunity guidelines. A search committee must be formed and charged with locating and recommending for hire the ideal candidate. In the past, our committees have struggled with where and how to advertise open positions; cost is of great concern, especially with many ongoing recruitments. Advertising in specialty or society journals (both print and Web) can be expensive, which is also the case with search firms.

The literature shows that the internet has a beneficial effect on job searches and, ultimately, successful employment. A 2012 study by Beard et al looked at whether internet use reduces job search costs, thus discouraging job seekers from giving up on active job searching and abandoning the labor market [3]. The study found that internet use has *significant positive effect on job search efforts* and actually reduces the chance that an unemployed individual will become disenchanted with job prospects and give up looking for a job altogether by 50%. This is likely because of the fact that one can search for jobs on the internet 7 days a week, 24 hours a day, and with only the minimal cost of internet access, if any. Job prospects are increased because geographic boundaries are minimized; one can search for employment in another city, state, or even country. Another study looking at young American job seekers found that those using the internet to find jobs had unemployment durations that were 25% shorter than those who looked for jobs solely offline [4]. Young Americans are increasingly searching the internet for employment. The Pew Research Center reported in 2011 that 24% of those aged younger than 40 years rely primarily on the internet for job information, whereas only 17% look to newspapers [5].

The internet presents unique opportunities for physician employment searches and can potentially overcome many of the logistical and financial barriers of traditional methodologies. In the medical field, where highly skilled individuals are sought for teaching and clinical positions, jobs advertisements were typically placed in printed journals, specialty and society-specific publications. Search (*headhunter*) firms were also employed, and many of those used—and still continue to use—print media to locate prospective employees.

Our historical experience as an employer was that there was a high, often prohibitive, cost associated with reaching wide audiences quickly. There was the fear that attractive potential candidates would miss time-limited advertisements in scant numbers of journals. Conversely, job seekers would risk missing our infrequent periodic or niche postings. The Web has dramatically altered that dynamic. As Frank and Taylor state, “The internet has offered ways to attack time, cost, and reach simultaneously” [6]. Employers want their positions to be found by search engines that identify low- or no-cost criteria and to be displayed on the first page or 2 of results. This is facilitated by the potential employee linking to the job description through an intermediary *aggregator*. In practical terms, it would be of the utmost importance for an employer to e-list a job in such a digital format to be picked up by a popular automated aggregator and hence then appear on the first page or 2 of search results.

Thus, our suggestion for employers with nonurgent position postings and limited budgets is to announce the job on their institution’s employment website (incurring essentially no cost). They then need to perform manual searches to check whether it was picked up by high-visibility aggregators (eg, Indeed). Suitable jobs for this approach might, for example, include primary care practitioners in large organizations with enough predicted attrition to justify always being on the look-out for new hires. If the posting is not detected by aggregators using their proprietary algorithms, then this might be accomplished by appropriately modifying the institution’s e-description or by paying the aggregator directly.

For employers with more pressing needs and larger budgets, we propose expanding searches by paying for either subscription journal listings or fee-based aggregators. Deciding between those 2 possibilities lacks clarity, especially as they have different inherent strengths, and the ultimate decider might be cost. For example, SubscrAg may also maintain a searchable database into which job seekers enter their qualifications and goals; however, employers would need the manpower to take advantage of those listings, otherwise the subscription could be a waste of money. In our case, we chose one of these subscription services (Practice Link) because we had personnel available to send (*blast*) emails to thousands of potential candidates who had expressed an interest in the specialty we were advertising. Reasons for instead choosing a SubscrJ include absence of such human resources, the market penetrance of a particular journal or its appeal to a certain type of medical position, or simply the cost. Budget allowing, we suggest advertising in both the SubscrAg and SubscrJ. The next step would be to expand the advertisements by including a fee-based specialty society posting, a SubscrSoc.

Finally, for urgent, failed, difficult (eg, limited candidate availability), or high-budget searches, serious consideration needs to be given to hiring a professional search firm. This can be much more expensive but does have its advantages: these companies typically maintain listings of attractive candidates, prescreen applicants (eg, digitally or by interview), identify an applicant pool by personally contacting leaders in the chosen field, target print mailings based on such parameters as specialty or location, and avoid the costs of large human resources departments.

The question arose during this study as to why certain sites are always in the top 10 and top 5. The answer has to do with the algorithms Google uses to determine its search results or relevancy rankings. These algorithms calculate which websites most accurately match users’ search terms [7]. For example, an individual searching for rheumatology jobs will likely find that the 5th search result will have more relevant job advertisements compared with the 19th result, which could be a low-budget portal with advertisements for mostly irrelevant jobs. Another question arose concerning why Indeed appears multiple times in the search results for a given specialty. On closer inspection, it appears that Indeed pulls job advertisements from multiple sites. According to Forbes, these *aggregator* sites, including Simply Hired, Indeed, Snagajob.com, and Beyond.com, *pull and reorganize postings from other job sites to make them easy to surf*, eliminating the need for job seekers to go to multiple corporate websites looking for job advertisements [8]. For example, our college of medicine had an open position for an endocrinologist, which it advertised on its own website, and yet the same advertisement appeared on Indeed. It is interesting to note that Monster and Career Builder, 2 sites synonymous with Web-based job advertisements, rarely appeared in the top 10 search results for any clinical specialty. These sites are clearly not the right forum for advertising or looking for clinical positions at academic institutions. It is also interesting to note that no search firms appeared in the top 5 results of any clinical specialty searched.

### Limitations of the Research

There are several known limitations to the research conducted. First, only 1 search engine, Google, was chosen to conduct the searches based solely on its popularity among the general public. A random search of 1 or 2 specialties with the chosen keywords using search engines Bing and Yahoo! did yield slightly different results as far as the order in which results appeared. This could affect the composition of the top 10 and top 5 search results. Timing is another limitation. Searches were conducted on different days and at different times over the course of several weeks. When the same search (eg, *endocrinology jobs*) was test searched on different days and times, the same overall results were obtained; however, the rank order of results varied slightly. This could affect the top 10 and top 5 search results, as it appears that Google's search results might have substantive changes in the approximate weekly time frame, but not daily. For example, someone searching for *endocrinology jobs* today may see Indeed as the tenth search result, but someone searching with the same terms next week may see Indeed at the ninth or even eleventh search result. However, after conducting 176 searches over the course of many weeks, the top 10 and top 5 search results rarely varied. Another concern regarding timing is that job seekers who are currently employed may be limited to searching only on weekends or in the evening hours; searches were not conducted during either of those 2 periods. Finally, our search approach only identifies potential candidates who are actively looking for jobs, as opposed to search firms that can expand the applicant pool to top talent not yet in the market to change employment. Exploring differences in expertise between active and passive candidates is outside the scope of this study.

## Conclusions

### Advice for Employers

On the basis of the results of these searches, the team conducting this study at the College of Medicine determined that the best approaches for future search committees charged with hiring clinicians at an academic institution are to (1) advertise on your institution's own website; (2) check if an NoCoAg picked up that advertisement from your website; (3) if not, alter your advertisement accordingly or pay to advertise there; (4) consider paying for what could be costly SubscrAg services, based on whether you have the staff and resources to take advantage of contacting prospective candidates who have registered and created a career profile that contains (albeit limited) data, which could narrow down the search; (5) pay to advertise with a SubscrJ, which will depend on the reputation and competitiveness of the various journals in the particular specialty; (6) pay to advertise on 1 society website, which may be less practical when a particular field has multiple such organizations.

### Advice for Job Seekers

On the basis of the results of these searches, job seekers should be aware of the following when searching for jobs on the internet: (1) trying multiple search engines might add a sense of completion but has diminishing returns; (2) test the search terms suggested in our tables but do not expect dramatically different results when viewing the first 2 search pages; (3) it is not likely to be productive to conduct the same search (ie, use the same keywords) multiple times per day, as you will likely get the same results—instead try again in a week's time; (4) pay attention to patterns in search results, particularly in the top 10—sites that keep appearing likely have more relevant content; (5) it may be beneficial to create a profile in a SubscrAg, as it appears so often in the top 10 and top 5, and employers who register have access to your information.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Top 10 Google search results for specialist and keyword jobs.

[PDF File (Adobe PDF File), 131KB - jmir_v21i7e12638_app1.pdf ]

## Multimedia Appendix 2

Top 5 Google search results for specialty and keyword jobs.

[PDF File (Adobe PDF File), 138KB - jmir_v21i7e12638_app2.pdf ]

## References

1. Purcell K, Brenner J, Raine L. Pew Research Center. 2012. Search Engine Use URL: http://www.pewinternet.org/2012/03/09/search-engine-use-2012/ [accessed 2015-07-01] [WebCite Cache ID 73PvT6HZn]
2. Jansen BJ, Spink A, Saracevic T. Real life, real users, and real needs: a study and analysis of user queries on the web. Inf Process Manag 2000;36(2):207-227. [doi: 10.1016/s0306-4573(99)00056-4]

3. Beard TR, Ford GS, Saba RP, Seals Jr RA. Internet use and job search. Telecomm Policy 2012;36(4):260-273. [doi: 10.1016/j.telpol.2011.12.001]
4. Kuhn P, Mansour H. Is internet job search still ineffective? Econ J 2014;124(581):1213-1233. [doi: 10.1111/ecoj.12119]
5. Rosenstiel T, Mitchell A, Purcell K, Raine L. Pew Research Center. 2011. Part 5: The Role of the Internet URL: http://www.pewinternet.org/2011/09/26/part-5-the-role-of-the-internet/ [accessed 2015-09-23] [WebCite Cache ID 73PviTc1y]
6. Frank FD, Taylor CR. Talent management: trends that will shape the future. Hum Resour Plan 2004;27(1):33-41.
7. Google Support. 2019. How Google Search Works URL: https://support.google.com/webmasters/answer/70897?hl=en [accessed 2015-07-01] [WebCite Cache ID 73Pvpvm35]
8. Marcus M. Forbes. 2009. The Best Way to Find (and Fill) a Job Online URL: http://www.forbes.com/2009/05/26/job-seeking-websites-entrepreneurs-human-resources-monster.html [accessed 2015-07-01] [WebCite Cache ID 73Pw2cjR0]

*Edited by G Eysenbach; submitted 29.10.18; peer-reviewed by G Claudine Mbemba; comments to author 25.04.19; revised version received 20.05.19; accepted 20.05.19; published 05.07.19.*

*Please cite as:*
*Gillum S, Williams N, Brink B, Ross E*
*Clinician Job Searches in the Internet Era: Internet-Based Study*
*J Med Internet Res 2019;21(7):e12638*
*URL: https://www.jmir.org/2019/7/e12638/*
*doi:10.2196/12638*
*PMID:31278735*

Original Paper

# Patient Attitudes About Viewing Their Radiology Images Online: Preintervention Survey

Ciarra Halaska[1*], MA; Peter Sachs[2*], MD; Kate Sanfilippo[1*], BSc; Chen-Tan Lin[3*], MD

[1]UCHealth, Aurora, CO, United States

[2]Department of Radiology, University of Colorado School of Medicine, Aurora, CO, United States

[3]University of Colorado School of Medicine, Aurora, CO, United States

[*]all authors contributed equally

**Corresponding Author:**
Peter Sachs, MD
Department of Radiology
University of Colorado School of Medicine
12401 E 17th Ave
Mail Stop L954
Aurora, CO,
United States
Phone: 1 720 848 4649
Email: peter.sachs@ucdenver.edu

## Abstract

**Background:** Although patient data is available through electronic portals, little information exists about the benefits and/or challenges of providing patients with online access to their radiology images.

**Objective:** The aims of this quality improvement project were to understand patient attitudes toward being able to view their radiology images online and determine how information should be presented to ensure the images are helpful to the patients, rather than causing confusion and anxiety.

**Methods:** An online survey of consumers was conducted to evaluate attitudes toward online access to personal radiological images.

**Results:** A total of 105 responses were received from 686 community members (15.3%). Of 105 consumers, 94 (89.5%) reported a desire to have access to the radiology images within their online patient portal; 86.7% (91/105) believed it would help them better understand their medical conditions and 81.0% (85/105) said this would help them feel more in control of their care. Most respondents (74/105, 70.5%) said it would help them feel reassured that their doctor was doing the right thing, and 63.8% (67/105) said it would increase their level of trust in their doctor. Among surveyed patients, 78.1% (82/105) valued viewing their radiology images online, while 92.4% (97/105) valued their online radiology reports. Most patients (69/105, 65.7%) wished to discuss their results with their ordering clinician, 29.5% (31/105) wished to discuss with their interpreting radiologist, and 3.8% (4/105) wished to share their images on social media. The biggest potential concern among 23.8% (25/105) was that the images would be confusing.

**Conclusions:** A large majority of surveyed patients desired the ability to view their radiology images online and anticipated many benefits and few risks. Health care organizations with electronic health records and online patient portals should consider augmenting their existing portals with this highly desired feature. To avoid the biggest patient concern, radiology reports should accompany images. Patients wanted to discuss their results with their ordering physician and their interpreting radiologist. Some even would like to share results on social media. Further research on the actual experience with such a tool will be needed.

***(J Med Internet Res 2019;21(7):e12595)*** doi:10.2196/12595

**KEYWORDS**

connected health; electronic health records; information transparency with patients; online patient-physician communication; online patient portal; radiology images; second opinion; social media; test result management

## Introduction

Online patient portals linked to electronic health records (EHRs) typically offer tools such as online communication between patient and clinician and patient access to portions of their medical record data, which includes test results, radiology reports, and pathology reports. Patients have long expressed a desire to view their medical reports, with expected benefits and few perceived risks [1]. More recently this includes viewing reports via online tools [2-6]. The nationwide Open Notes initiative took this further in 2011 [7,8]. Surveys demonstrated patient interest in viewing their radiology reports and images, and patients perceive there are potential benefits from doing so [9,10]. Although other institutions, such as the Mayo Clinic and the Department of Veterans Affairs have image-enabled their patient portals, to our knowledge there have been few published surveys regarding patient perceptions of the advantages of being able to view their own radiology images online. Greco et al [11] conducted a limited survey of patients viewing images in a personal health record (PHR) and found they responded favorably to having access to their images while expressing a high level of concern for the privacy of their health information.

A group composed of clinical and imaging informatics specialists focused on improving health care technology developed a method to offer patients the ability to view radiology images associated with their radiology report in the online patient portal. However, we were uncertain about how this new tool might be received by the patient population. Therefore, we decided to solicit opinions from an online community in order to gauge interest prior to implementation.

## Methods

### Survey Design

We designed a survey to evaluate attitudes of patients regarding the ability to view their own radiological images online; it was intended to serve as a baseline survey to be refielded postimplementation (see Multimedia Appendix 1 for the survey instrument). The project was initiated by an information technology program manager (KS), senior market research analyst (CH), physician imaging informaticist (PS), and chief medical information officer (C-TL). We conducted this project as market research and quality improvement with an established panel of volunteer community members across Colorado; therefore, the study was exempt from review by the institutional review board.

### Target Population

This was a preintervention survey of 686 community members, many of whom were patients at UCHealth. The UCHealth Insights Community exists to collect community feedback in order to help UCHealth elevate how health care providers interact with the community and continuously evolve the health care experience and reflects the demographics of the health care decision makers in Colorado. Members are recruited via community partnerships, social media, and the UCHealth website, including My Health Connection (UCHealth's online patient portal). The screening criteria required for community members to join the Insights Community are that they must live in Colorado, cannot work in health care, must be involved in their household's health care decisions, must have health insurance, and cannot participate in more than three online research communities. There were no additional participation criteria set for this quality improvement project. This online feedback community is hosted by a commercial entity (My-Take.com).

### Recruitment

Potential participants received one email invitation describing the purpose of the survey and the opportunity to participate and one email reminder if they had not yet completed the survey. With previous surveys in this community, further reminders did not increase response rates. The survey was available for six days. Participants were reminded that their participation would earn them an entry into a monthly prize drawing, in accordance with the UCHealth Insights Community site policy. This prize drawing was standard procedure for the Insights Community independent of this particular survey. Participants from the Insights Community were surveyed in July 2018.

### Statistical Analysis

Descriptive information regarding demographics of participants was collected along with behavioral components related to this survey—specifically whether they were a UCHealth patient, had used radiology services in the past year, and had used an online patient portal at that time to manage their health. To gauge interest in the option to have radiology images provided online within their patient portal, we examined the distribution of responses to questions across a range of categorical and ordinal variables within the survey. After reading through all responses, open-ended responses were coded manually, sorting each response into a bucket. Buckets were developed by determining which themes came up most frequently. One response could be coded into multiple buckets, for example "Understanding the results and why my doctor is making recommendations. Sharing with surgeons, etc" was coded as both "increase understanding" and "conveniently see/share results." We rated responses to quantitative questions about value from 1=not at all valuable to 5=extremely valuable. We rated responses to questions about level of agreement from 1=strongly disagree to 5=strongly agree. In order to report the survey results as percentages, Likert scores are reflected as top 2 box scores or scores of a 4 or 5 on these 5-point scales. We conducted all statistical analyses using Q Research Software version 5.4.5.0 (Displayr).

## Results

### Target Population

We received a total of 105 responses from a target population of 686 members who were invited to participate in this research, yielding a 15.3% response rate. The mean age of the sample was 50.37 (SD 15.08) years, and most respondents were female (74/105, 70.5%). A majority of the sample identified as white (85/105, 81.0). All respondents lived in Colorado, and 59.0 (62/105) of respondents were UCHealth patients. Over half (60/105, 57.1%) had used radiology services in the past year, and approximately half of those respondents (31/60, 52%) had viewed their radiology report online. Most participants (80/105, 76.2%) used their online patient portal regardless of where they received care. The sample size was too small to detect differences in responses from those with and without portal accounts. There were no significant demographic differences between responders and nonresponders.

### Survey Results

When asked whether they would want access to future radiology images online, 89.5% (94/105) said yes. Of these 94 respondents, 33 (35%) explained in the open-ended follow-up that this would help them to understand their radiology results at a higher level, especially when paired with their radiology report, and 27 respondents (29%) reported they desired the convenience of having access to their images online; 89.5% (94/105) of responders liked the idea of viewing their images in the online patient portal. The 10.5% (11/105) who disliked the idea were asked why. Open-ended responses indicated that people did not see a need for this offering (3/11) and would rather view and discuss their images in person with their doctor (4/11).

When assessing the value of online access to radiology images compared to radiology reports, 78.1% (82/105) found value in having online access to radiology images compared with 92.4% (97/105) who found value in radiology reports.

Participants rated their level of agreement with various statements regarding online access to radiology images. The full distribution can be found in Table 1. Access to online radiology images received high top 2 box scores for improved patient-doctor relationships: increased levels of trust (67/105, 63.8%), feeling reassured that their doctor is doing the right thing (74/105, 70.5%), and being able to better follow their doctor's recommendations (74/105, 70.5%). Participants also agreed that viewing their images would help them to understand their medical conditions (91/105, 86.7%) and feel more in control of their health care (85/105, 81.0%). Few participants reported they were concerned with finding errors (10/105, 9.5%) or being worried (10/105, 9.5%) or confused (9/105, 8.6%) when viewing their images.

Anticipating that patients would have questions about their images, we asked participants how they would use their online images, and with whom they would most like to discuss their images; responses are shown in Table 2. Of the respondents, 80.0% (84/105) reported they would like the ability to share their images with their doctor, 78.1% (82/105) desired to save a copy directly from their online portal, 62.9% (66/105) would share their images with other doctors for a second opinion, and 3.8% (4/105) of our respondents indicated they would be interested in sharing their radiology images on social media.

Results to the voluntary open-ended question regarding the benefits of viewing images online are shown in Table 3. The convenience of seeing, saving, and sharing their images was cited by 48.6% of respondents (51/105); 46.7% (49/105) indicated that it would increase their understanding.

> *It would help me understand my condition better, and it would give me more time to formulate questions for my doctor.*

**Table 1.** Distribution of agreement with statements regarding online viewing of images (N=105).

| Statement: I believe that viewing my radiology images online would cause me to... | Agreement with statement, scored on a 5-point scale, n (%) | | |
|---|---|---|---|
| | Disagree (bottom 2) | Neutral (middle 3) | Agree (top 2) |
| ...feel confused or have a lot of questions. | 70 (66.7) | 26 (24.8) | 9 (8.6) |
| ...worry more. | 74 (70.5) | 21 (20.0) | 10 (9.5) |
| ...find errors in my radiology reports. | 68 (64.8) | 27 (25.7) | 10 (9.5) |
| ...trust my doctors more. | 15 (14.3) | 23 (21.9) | 67 (63.8) |
| ...feel reassured. | 8 (7.6) | 23 (21.9) | 74 (70.5) |
| ...better follow recommendations. | 13 (12.4) | 18 (17.1) | 74 (70.5) |
| ...feel more in control. | 6 (5.7) | 14 (13.3) | 85 (81.0) |
| ...better understand my medical condition. | 4 (3.8) | 10 (9.5) | 91 (86.7) |

**Table 2.** Participant responses to the question: What would you do with your online images? (N=105)

| Response | Agreement with statement, n (%) |
|---|---|
| Share them with my primary care doctor, if they don't have them already | 84 (80.0) |
| Save a copy for my records | 82 (78.1) |
| Share them with other doctors for a potential second opinion | 66 (62.9) |
| Share them on social media | 4 (3.8) |
| Other | 8 (7.6) |
| None of the above | 6 (5.7) |

**Table 3.** Benefits described by participants when asked the open-ended question: What would the benefits be of viewing your radiology images online? (N=105)

| Benefit of viewing images | Value n (%) |
|---|---|
| Conveniently see, save, and share images | 51 (48.6) |
| Increase understanding | 49 (46.7) |
| Increase level of control | 12 (11.4) |
| Use for second opinion from another doctor | 10 (9.5) |
| Formulate/ask questions | 9 (8.6) |
| Cool/curiosity | 7 (6.7) |
| Avoid follow-up appointment | 3 (2.9) |

**Table 4.** Participant concerns expressed in response to the open-ended statement: Please explain any concerns about viewing your radiology images online. (N=105)

| Concern | Value n (%) |
|---|---|
| No concerns | 50 (47.6) |
| Would cause confusion | 25 (23.8) |
| Security/privacy concerns | 19 (18.1) |
| Can’t ask questions or get answers quickly | 19 (18.1) |
| Mistakes (wrong images could be uploaded) | 3 (2.9) |

Participants also expressed how tedious it currently is to obtain a copy of their radiology images.

> *It saves me the hassle of filling out a form, sending an email, making a trip to the doctor’s office, making four phone calls, and spending an hour on hold to make sure my specialist has everything they need for my appointment. That is what I had to do for the images, and the report still didn't come across, so when I had a few moments I logged into the portal and printed out the report to bring with me. I would much rather take the self-service route to save me the time and stress of getting images from the facility to the doctor, plus I would have the opportunity to see the images myself and better understand the issue when discussing it with my doctor.*

Participants concerns when asked the voluntary open-ended question about viewing their images online are shown in Table 4. Of the respondents, 47.6% (50/105) said they had no concerns. Concerns regarding confusion (23.8%, 25/105) and security/privacy (18.1%, 19/105) were the most prevalent.

> *Not understanding what I'm looking at is a big [concern]. There is a reason that doctors go to school to learn how to read images. This is why the radiology report is also essential. It ties the two together.*

> *I worry a little about security but would hope that you guys provide encrypted portals, passwords, etc.*

In response to question 11, 65.7% of respondents (69/105) reported they would prefer to discuss their images with the doctor who referred them to radiology, and 29.5% (31/105) would rather discuss with the radiologist who wrote their report.

The final question of the survey asked participants to rate the idea of being able to view their radiology images in their online patient portal. The response from 86.7% (91/105) was favorable, which resulted in a top 2 box score.

## Discussion

### Principal Findings

Our main finding is that our respondents had a high level of interest in viewing their images with a surprisingly low level of concern about possible confusion. Overall, patients preferred to view their images alongside the reports, as viewing images without the report was seen as potentially confusing. Patients also had a strong interest in being able to download their images from the portal. Patients expressed frustration with the current process that requires they come to the radiology file room, request a CD of images, then physically transport the CD to the recipient, who may not have a CD drive. A small percentage of patients also expressed an interest in sharing their images on social media.

Multiple previous reports, including the Open Notes literature [7,8], have described the benefits of patient access to their medical record via an EHR-based patient portal in both the ambulatory and inpatient settings. These include a sense of increased feelings of empowerment and autonomy, “including control, understanding, reassurance, and following recommendations” [12,13]. Patients have also described feeling an increased ability to coordinate their own care and remember important health care tasks and an improved sense of participating in their own care [4]. Anticipated risks such as increased patient worry and increased provider workload have not been borne out [12]. Interestingly, such access has not been shown to improve health status [3].

Radiology has also embraced the value to both patients and the specialty of increased visibility of imaging reports to and greater interaction with patients, with the Radiology 3.0 initiative serving as a foundation [14]. This has focused on direct contact with patients via consultation clinics or widespread face-to-face transmission of reports to patients by interpreting radiologists [15]. To date, this has gained the most traction in the breast imaging sphere [16]. Patients have indicated their desire to view their reports quickly and in understandable detail [5]. Multiple reports have demonstrated that patients value online access to their imaging reports and that doing so does not increase workloads for the ordering providers [17]. Interestingly, patients have indicated that although they value direct access to their reports, they prefer to first hear the results from their ordering provider and prefer to review their images with their provider when receiving the results [9,18]. While some patients do not wish to view their images, for those who do the experience can offer a powerful way to connect with their illness, with their provider, and should be offered [10].

### Limitations

This baseline survey had a low response rate (15%) among community members, consisted mostly of female respondents (70%), and was performed at one institution in one geographical location, which may limit the generalizability of these findings. However, the response rate was typical of response rates in other market research surveys. For example, Greco et al [11] had a 19.6% response rate in their survey of patients using a PHR for image exchange and viewing. In question 6, we asked about the value of viewing reports and viewing images and did not give an option for viewing images with reports. This may have been confusing to respondents and makes interpretation of the responses potentially ambiguous. Our community members expressed a high level of interest in radiology images, which may reflect members who were patient portal users (76%). However, the level of high interest is similar to the greater than 80% interest in images seen in previous reports [9].

### Comparison With Prior Work

A consortium of four academic institutions (Mt. Sinai New York, University of California San Francisco, Mayo Clinic, and University of Maryland) collaborated to provide their patients with a personal heath record (PHR) that included image viewing using the Radiological Society of North America's Image Share Network [11]. We note that, unlike our integrated project, this publicly available image viewing site is separate from their health care organization's patient portal. Their survey of users focused primarily on satisfaction with the specific technologies available (eg, CD or internet) for image downloading and viewing. They found that 96.5% responded favorably to having access to their images and imaging reports and 78% viewed their images independent of their providers. Patients were concerned about the privacy of the PHR (67% to 78%), but other potential concerns were not explored. They also did not assess patient preference for viewing their images with others (provider, family, social media) nor did they solicit additional feedback as we did.

### Future Directions

Our health system is in the process of implementing radiology image viewing through the online patient portal and plans to report on the findings and experiences of patients actually viewing their images online. We will enable image viewing using the currently available enterprise viewer provided with our Picture and Archiving Communication System. This is a view-only system currently optimized for desktop viewing. Based on the survey feedback, patients will be able to download and/or share their images using an existing commercial image exchange tool, which is currently used by the health system for image exchange between health care institutions and providers but not yet embedded in the patient portal. This functionality will allow patients to upload and download their images via CD, thumb drive, and the cloud. The enterprise is also currently evaluating mobile friendly viewers that will further enhance and simplify the patient experience. An important means for supporting patient's questions about radiology images will be a tool allowing easy communication between the patient and the ordering provider as well as the interpreting radiologist (personal communication by Alexander J. Towbin, MD, on 10/28/2018). As momentum grows for immediate access by patients to all their images and reports, the ability to communicate directly with the interpreting radiologist may alleviate some ordering providers concern about potential patient anxiety and possible increased burden on them to respond to questions [19,20].

The interest in sharing radiology images on social media is worthy of exploration in future work. Possible motivations include ultimate transparency (eg, "I want everyone to know what I have"), seeking a second opinion (eg, "Maybe if I tweet this, some radiologists somewhere will have a different idea"), finding a community (eg, "Maybe someone else out there has the same thing and we can be friends"), or perhaps a cool factor (ie, "No one else can do this, look at me!").

### Conclusions

A large majority of surveyed patients desired the ability to view their radiology images online and anticipated many benefits and few risks. Health care organizations with EHRs and online patient portals should consider augmenting their existing portals with this highly desired feature. Further research on actual experience with such a tool will be needed and helpful.

## Authors' Contributions

CH, KS, PS, and C-TL codesigned the survey and cowrote the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Preintervention survey.

[PDF File (Adobe PDF File), 85KB - jmir_v21i7e12595_app1.pdf ]

## References


1. Ross SE, Lin C. 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]
2. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin C. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
3. Earnest MA, Ross SE, Wittevrongel L, Moore LA, Lin C. Use of a patient-accessible electronic medical record in a practice for congestive heart failure: patient and physician experiences. J Am Med Inform Assoc 2004;11(5):410-417 [FREE Full text] [doi: 10.1197/jamia.M1479] [Medline: 15187074]
4. Ross SE, Todd J, Moore LA, Beaty BL, Wittevrongel L, Lin C. Expectations of patients and physicians regarding patient-accessible medical records. J Med Internet Res 2005;7(2):e13 [FREE Full text] [doi: 10.2196/jmir.7.2.e13] [Medline: 15914460]
5. Johnson AJ, Easterling D, Williams LS, Glover S, Frankel RM. Insight from patients for radiologists: improving our reporting systems. J Am Coll Radiol 2009 Nov;6(11):786-794. [doi: 10.1016/j.jacr.2009.07.010] [Medline: 19878886]
6. Lee CI, Langlotz CP, Elmore JG. Implications of direct patient online access to radiology reports through patient web portals. J Am Coll Radiol 2016 Dec;13(12 Pt B):1608-1614. [doi: 10.1016/j.jacr.2016.09.007] [Medline: 27888949]
7. 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 01;24(e1):e166-e172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
8. Crotty BH, Anselmo M, Clarke D, Elmore JG, Famiglio LM, Fossa A, et al. Open Notes in teaching clinics: a multisite survey of residents to identify anticipated attitudes and guidance for programs. J Grad Med Educ 2018 Jun;10(3):292-300 [FREE Full text] [doi: 10.4300/JGME-D-17-00486.1] [Medline: 29946386]
9. Cabarrus M, Naeger DM, Rybkin A, Qayyum A. Patients prefer results from the ordering provider and access to their radiology reports. J Am Coll Radiol 2015 Jun;12(6):556-562. [doi: 10.1016/j.jacr.2014.12.009] [Medline: 25892226]
10. Carlin LE, Smith HE, Henwood F. To see or not to see: a qualitative interview study of patients' views on their own diagnostic images. BMJ Open 2014 Jul 31;4(7):e004999 [FREE Full text] [doi: 10.1136/bmjopen-2014-004999] [Medline: 25082418]
11. Greco G, Patel AS, Lewis SC, Shi W, Rasul R, Torosyan M, et al. Patient-directed internet-based medical image exchange: experience from an initial multicenter implementation. Acad Radiol 2016 Feb;23(2):237-244 [FREE Full text] [doi: 10.1016/j.acra.2015.10.012] [Medline: 26625706]
12. Pell JM, Mancuso M, Limon S, Oman K, Lin C. Patient access to electronic health records during hospitalization. JAMA Intern Med 2015 May;175(5):856-858. [doi: 10.1001/jamainternmed.2015.121] [Medline: 25751393]
13. Weitzman ER, Kaci L, Mandl KD. Acceptability of a personally controlled health record in a community-based setting: implications for policy and design. J Med Internet Res 2009;11(2):e14 [FREE Full text] [doi: 10.2196/jmir.1187] [Medline: 19403467]
14. Ellenbogen PH. Imaging 3.0: what is it? J Am Coll Radiol 2013 Apr;10(4):229. [doi: 10.1016/j.jacr.2013.02.011] [Medline: 23545079]
15. Gunn AJ, Mangano MD, Choy G, Sahani DV. Rethinking the role of the radiologist: enhancing visibility through both traditional and nontraditional reporting practices. Radiographics 2015;35(2):416-423. [doi: 10.1148/rg.352140042] [Medline: 25763726]
16. Itri JN. Patient-centered radiology. Radiographics 2015 Oct;35(6):1835-1846. [doi: 10.1148/rg.2015150110] [Medline: 26466190]
17. Henshaw D, Okawa G, Ching K, Garrido T, Qian H, Tsai J. Access to radiology reports via an online patient portal: experiences of referring physicians and patients. J Am Coll Radiol 2015 Jun;12(6):582-586. [doi: 10.1016/j.jacr.2015.01.015] [Medline: 26047399]
18. Kuhlman M, Meyer M, Krupinski EA. Direct reporting of results to patients: the future of radiology? Acad Radiol 2012 Jun;19(6):646-650. [doi: 10.1016/j.acra.2012.02.020] [Medline: 22578223]
19. Woolen S, Kazerooni EA, Wall A, Parent K, Cahalan S, Alameddine M, et al. Waiting for radiology test results: patient expectations and emotional disutility. J Am Coll Radiol 2018 Feb;15(2):274-281. [doi: 10.1016/j.jacr.2017.09.017] [Medline: 29132998]
20. Gefen R, Bruno MA, Abujudeh HH. Online portals: gateway to patient-centered radiology. AJR Am J Roentgenol 2017 Nov;209(5):987-991. [doi: 10.2214/AJR.17.18291] [Medline: 28726503]

## Abbreviations

**EHR:** electronic health record
**PHR:** personal health record

*Edited by G Eysenbach; submitted 25.10.18; peer-reviewed by S Buttigieg, F Bourgeois, D Klein, S Chen, J Vidal-Alaball; comments to author 01.04.19; revised version received 12.05.19; accepted 18.06.19; published 18.07.19.*

*Please cite as:*
*Halaska C, Sachs P, Sanfilippo K, Lin CT*
*Patient Attitudes About Viewing Their Radiology Images Online: Preintervention Survey*
*J Med Internet Res 2019;21(7):e12595*
*URL: http://www.jmir.org/2019/7/e12595/*
*doi:10.2196/12595*
*PMID:31322124*

Original Paper

# Acute Care Patient Portal Intervention: Portal Use and Patient Activation

Kumiko O Schnock[1,2], RN, PhD; Julia E Snyder[1], BSc; Theresa E Fuller[1], BSc; Megan Duckworth[1], BA; Maxwell Grant[1], BA; Catherine Yoon[1], MS; Stuart Lipsitz[1,2], ScD; Anuj K Dalal[1,2], MD; David W Bates[1,2], MD, MSc; Patricia C Dykes[1,2], RN, PhD

[1]Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, United States

[2]Harvard Medical School, Boston, MA, United States

**Corresponding Author:**
Kumiko O Schnock, RN, PhD
Division of General Internal Medicine and Primary Care
Brigham and Women's Hospital
OBC-3rd Floor
1620 Tremont Street
Boston, MA, 02120-1613
United States
Phone: 1 617 525 8898
Fax: 1 617 732 7072
Email: kschnock@bwh.harvard.edu

## Abstract

**Background:** Patient-facing health information technology (HIT) tools, such as patient portals, are recognized as a potential mechanism to facilitate patient engagement and patient-centered care, yet the use of these tools remains limited in the hospital setting. Although research in this area is growing, it is unclear how the use of acute care patient portals might affect outcomes, such as patient activation.

**Objective:** The aim of this study was to describe the use of an acute care patient portal and investigate its association with patient and care partner activation in the hospital setting.

**Methods:** We implemented an acute care patient portal on 6 acute care units over an 18-month period. We investigated the characteristics of the users (patients and their care partners) of the patient portal, as well as their use of the portal. This included the number of visits to each page, the number of days used, the length of the user's access period, and the average percent of days used during the access period. Patient and care partner activation was assessed using the short form of the patient activation measure (PAM-13) and the caregiver patient activation measure (CG-PAM). Comparisons of the activation scores were performed using propensity weighting and robust weighted linear regression.

**Results:** Of the 2974 randomly sampled patients, 59.01% (1755/2974) agreed to use the acute care patient portal. Acute care patient portal enrollees were younger, less sick, less likely to have Medicare as their insurer, and more likely to use the Partners Healthcare enterprise ambulatory patient portal (Patient Gateway). The most used features of the acute care patient portal were the laboratory test results, care team information, and medication list. Most users accessed the portal between 1 to 4 days during their hospitalization, and the average number of days used (logged in at least once per day) was 1.8 days. On average, users accessed the portal 42.69% of the hospital days during which it was available. There was significant association with patient activation on the neurology service ($P<.001$) and medicine service ($P=.01$), after the introduction of HIT tools and the acute care patient portal, but not on the oncology service.

**Conclusions:** Portal users most often accessed the portal to view their clinical information, though portal usage was limited to only the first few days of enrollment. We found an association between the use of the portal and HIT tools with improved levels of patient activation. These tools may help facilitate patient engagement and improve outcomes when fully utilized by patients and care partners. Future study should leverage usage metrics to describe portal use and assess the impact of HIT tools on specific outcome measures in the hospital setting.

***(J Med Internet Res 2019;21(7):e13336)*** doi:10.2196/13336

**KEYWORDS**

patient portals; patient participation; patient activation; patient-centered care; inpatients

## Introduction

### Background

The acute care setting presents challenges for patients and their care partners who often feel disengaged and disempowered [1]. The experience can be isolating and uncertain, and patients are often left out of the decision-making process [2]. Engaging patients and encouraging active participation in their care may help address these issues and has the potential to improve health outcomes as well as the quality and safety of care [3]. Health information technology (HIT) has been shown to promote patient engagement and patient-centered care [4]. Previous research by our group found that engaging patients and health care providers in the intensive care unit using patient-centered HIT tools was associated with a reduction in adverse events and improved patient satisfaction [5].

Providing patients access to their personal health records and health care information through patient portals may improve patient satisfaction, outcomes, and safety [6-9]. Given the incentives associated with providing patients access to health information through the Meaningful Use program, outpatient portals are becoming increasingly common [10]; however, the use of patient portals during acute care hospitalizations remains limited [11]. Although research on acute care patient portals is expanding [7,9,12,13], few large-scale clinical trials have been conducted, and evidence supporting their impact on improved health outcomes is currently insufficient. [4,11-15].

Patient activation represents an important outcome measure. It refers to a patient's knowledge, skills, and confidence in managing their health condition [15]. Patient activation can be an indicator of patient engagement [16]. High levels of patient activation have been associated with lower costs and better outcomes [17-19]. Patient portals may represent a mechanism to improve patient activation; however, there is limited research assessing their association with patient activation in the acute care setting. For example, a recent randomized controlled trial conducted by Masterson-Creber et al found that access to an acute care patient portal did not significantly improve patient activation [20]. Similarly, O'Leary found that the use of an acute care patient portal had no significant effect on patient activation scores [21]. Another study found that patient activation scores increased over time with the use of an acute care patient portal designed for patients undergoing hematopoietic cell transplantation, but not linearly, suggesting that a *sweet spot* of utilization may exist [22]. However, these previous studies had small sample sizes and the study results may be insufficient to characterize the association of the use of an acute care patient portal with patient activation.

### Purpose of the Study

In this study, we conducted a large-scale intervention, implementing a patient portal, along with a suite of patient and provider-facing tools, to promote patient-centered care in the acute care setting. We assessed portal usage and analyzed the association between the acute care portal and patient activation. We hypothesized that successful implementation and use of the patient portal by inpatients would result in greater knowledge of their care and increased patient activation.

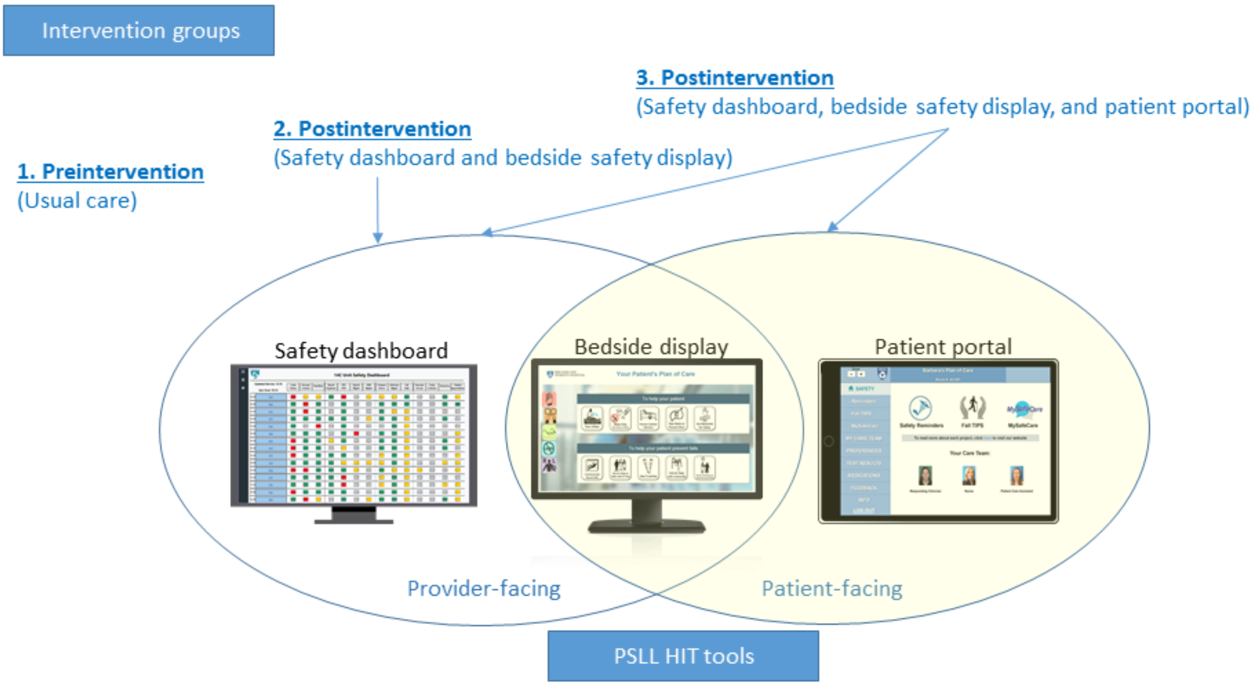
## Methods

### Setting and Participants

The patient portal was developed and implemented as part of an Agency for Healthcare Research and Quality (AHRQ)–funded Patient Safety Learning Laboratory (PSLL) project at Brigham and Women's Hospital (BWH), a large tertiary care center in Boston, Massachusetts. The PSLL project aimed to develop and implement a suite of HIT tools to engage patients and providers in improving quality and safety in the acute care setting. The HIT tools included a provider-facing safety dashboard [23], bedside safety display [24], and a patient portal (Figure 1). The provider-facing safety dashboard was used as a care team rounding tool and was only accessed by the health care providers [23]. The bedside safety display was both provider and patient facing. All patients were continuously exposed to the personalized bedside display monitor, whereas providers were intermittently exposed when in the patient room [24]. Randomly selected patients also received the patient portal that was a patient-facing tool accessed via a tablet computer or mobile device. These tools were implemented for an 18-month period from December 2016 to May 2018.

The PSLL tools were implemented using a randomized stepped-wedge design. Implementation of the intervention and recruitment for the patient portal began on a new inpatient unit every 1 to 2 months. In total, 6 units participated, including 3 general medicine, 1 neurology, and 2 oncology units. Patient activation was measured at 2 time points with 3 distinct groups of patients: (1) preintervention (usual care), (2) postintervention (patients exposed to the safety dashboard and bedside display), and (3) postintervention (patients exposed to the safety dashboard, bedside display, and patient portal).

**Figure 1.** Patient safety learning laboratory and health information technology tools and intervention groups. PSLL: Patient Safety Learning Laboratory; HIT: health information technology.




### The Patient Portal

The patient portal was a Web-based application specifically designed for the acute care setting. It leveraged vendor-based (Epic Systems Inc) electronic health record (EHR) data to provide patients and families access to the real-time information and educational content needed to proactively engage in their care during hospitalization. Features of the portal included the following: personalized safety reminders and a fall prevention plan (Fall TIPS) [25]; names and photos of care team members, medication lists, laboratory test results; a method to report safety concerns (MySafeCare) [26]; and general hospital information (Multimedia Appendix 1). In December 2017, a discharge preparedness checklist was added to the portal, and in March 2018, a safety issues dashboard was added to enhance the portal content and promote patient engagement. A mobile app was also developed with the same functions and features. The user interface and content of the patient portal were developed in collaboration with patients and their care partners through participatory iterative design [5,24].

### Recruitment and Enrollment

Research staff approached randomly selected patients or their care partners (health care proxies) on study units each weekday to offer the use of the patient portal. Patients who did not speak English, were not alert and oriented, had impairments that prohibited the use of the portal, or did not have a health care proxy were excluded. All other patients on the study units and on a medicine, neurology, or oncology service were eligible to participate. Patients were offered use of the portal on a tablet computer (iPad; Apple, Inc) provided by the study or their own device for the duration of their hospital stay. The mobile version was offered beginning in December 2017, available to download as a mobile app on Apple devices. An email address (or username) was required to set up a secure account and the research staff gave a brief orientation to the portal. More than one user could be created with the patient's permission (eg, patient and family member). The study staff provided their contact information, including an email address and phone number, for additional support. All study activities were approved by the Partners Healthcare Institutional Review Board.

### Measures and Data Analysis

We measured patient and care partner use of the portal by recording user actions in our database and leveraged previously reported measures of portal usage for comparison [7,11,20,22,27-29]. Measures included the number of visits to each page, the number of days used, length of users' access period, and average percent of days used during the access period. Demographic characteristics of patients who enrolled and patients who declined to participate were obtained from our EHR, and differences in portal users and nonusers were compared using a Fischer exact test and robust chi-square tests [30]. Owing to a technical issue in our database, portal activity for 136 users (136/1755, 7.75%) was not recorded. We compared the patient characteristics of this group with the other enrollees' usage data in Table 1. There were no significant differences between the 2 groups; therefore, we conducted the analysis without the missing data.

**Table 1.** Patient characteristics.

| Variable | Patients that enrolled to portal (n=1755) | | | Patients that declined portal (n=1219) | *P* value |
|---|---|---|---|---|---|
| | Enrolled with usage data | Enrolled without usage data | *P* value | | |
| Number of unique patients, n (%) | 1619 (92.25) | 136 (7.74) | —[c] | 1219 (100.00) | — |
| Age (years), mean (SD) | 56.84 (17.29) | 54.96 (19.00) | .35 | 61.21 (16.69) | <.001 |
| Female, n (%) | 883 (54.54) | 81 (59.56) | .26 | 646 (52.99) | .30 |
| **Race, n (%)** | | | | | |
| White | 1256 (77.58) | 103 (75.74) | .58 | 948 (77.77) | .52 |
| Black or African American | 202 (12.48) | 16 (11.76) | — | 162 (13.29) | — |
| Asian | 28 (1.73) | 4 (2.94) | — | 23 (1.89) | — |
| Other[a] | 101 (6.24) | 8 (5.88) | — | 68 (5.58) | — |
| Unavailable | 21 (1.30) | 2 (1.47) | — | 10 (0.82) | — |
| Declined | 11 (0.68) | 3 (2.21) | — | 8 (0.66) | — |
| **Ethnicity, n (%)** | | | | | |
| Hispanic or Latino | 111 (6.86) | 8 (5.88) | .91 | 67 (5.50) | .34 |
| Non-Hispanic | 1461 (90.24) | 124 (91.18) | — | 1119 (91.80) | — |
| Unavailable | 47 (2.90) | 4 (2.94) | — | 32 (2.71) | — |
| **Primary language, n (%)** | | | | | |
| English | 1525 (94.19) | 128 (94.12) | .95 | 1158 (95.00) | .16 |
| Spanish | 26 (1.61) | 2 (1.47) | — | 10 (0.82) | — |
| Other | 22 (1.36) | 2 (1.47) | — | 11 (0.90) | — |
| Unavailable | 46 (2.84) | 4 (2.94) | — | 40 (3.28) | — |
| Charlson score, mean (SD) | 2.39 (2.80) | 2.26 (2.66) | .76 | 3.23 (3.02) | <.001 |
| **Insurance, n (%)** | | | | | |
| Private | 777 (47.99) | 73 (53.68) | .63 | 498 (40.85) | <.001 |
| Medicaid | 161 (9.94) | 9 (6.62) | — | 106 (8.70) | — |
| Medicare | 626 (38.67) | 49 (36.03) | — | 572 (46.92) | — |
| Self-pay | 36 (2.22) | 3 (2.21) | — | 27 (2.21) | — |
| Other | 19 (1.17) | 2 (1.47) | — | 16 (1.31) | — |
| Median income by zip code, mean (SD) | 71,165.68 (26,541.36) | 72,658.02 (28,874.67) | .56 | 72,054.16 (27,478.91) | .44 |
| Patient Gateway users, n (%) | 286 (17.67) | 32 (23.53) | .09 | 168 (13.78) | .002 |
| Length of stay[b], mean (SD) | 8.85 (8.99) | 5.83 (5.83) | <.001 | 8.61 (8.44) | .63 |

[a]Includes *Hispanic or Latino* as a race choice.

[b]First admission to time of first enrollment or decline.

[c]No data.

We used the short form of the patient activation measure (PAM-13) [31] to assess patient activation. For patients who could not participate in the PAM-13 survey, we surveyed their caregivers using the caregiver version of PAM-13 (caregiver patient activation measure [CG-PAM]) [32] to assess their activation. The PAM-13 and CG-PAM are validated 13-item instruments, with scores ranging from 0 to 100, measuring patient skill, knowledge, and confidence for self-management of health conditions [31,32]. The PAM-13 has been shown to be reliable in both outpatient and inpatient settings [33]. The PAM-13 (or CG-PAM) was administered to a random sample of patients—including both patient portal users and nonportal users—across all study units before and after the implementation of the PSLL intervention. Research staff approached patients at multiple times throughout the day using randomized lists. In all cases, patients were approached approximately one to two days before their discharge. PAM surveys were anonymous and only identified on the unit level.

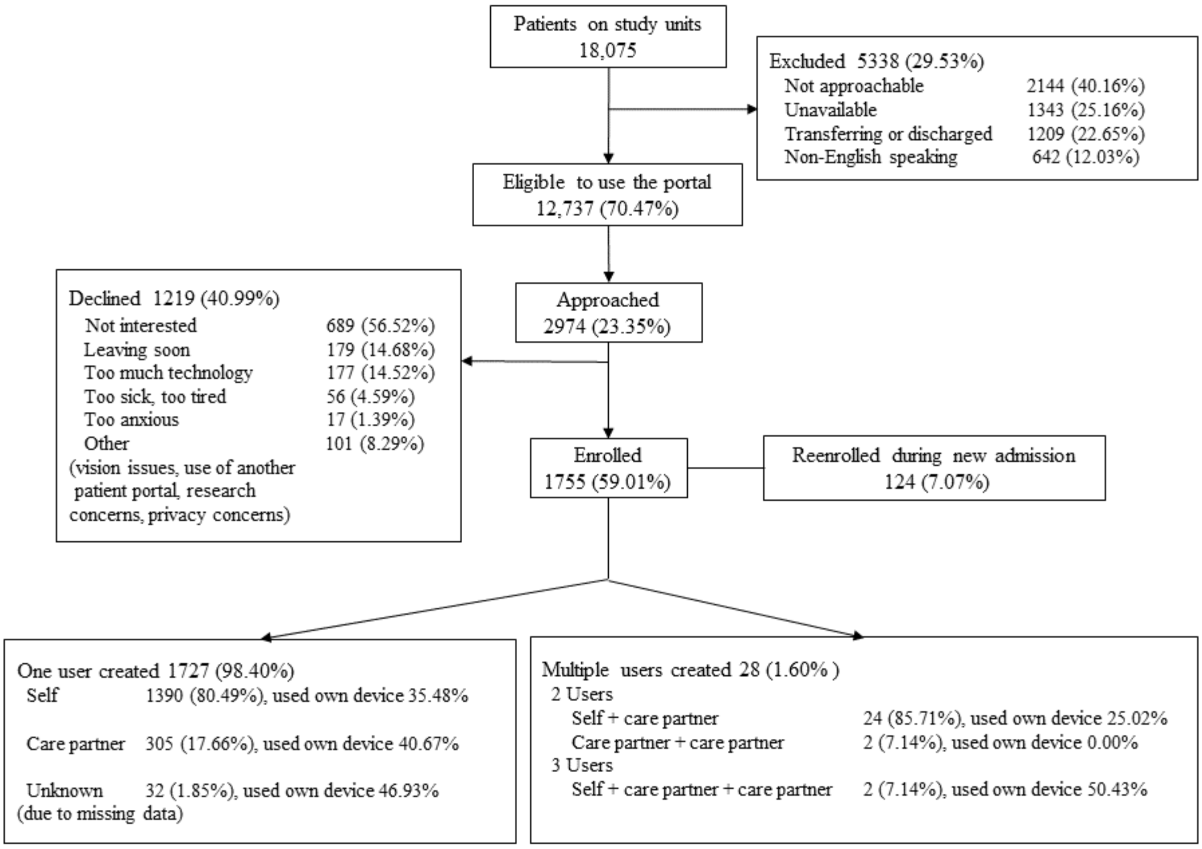
The preintervention period began 3 months before the implementation of the intervention (September-November 2016), and the postintervention period occurred for 18 months after the first implementation (December 2016-May 2018). Within each of the 3 services (medicine, oncology, and neurology), we compared PAM scores across the 3 groups (see Figure 1): (1) preintervention (usual care), (2) postintervention (safety dashboard and bedside display), and (3) postintervention (safety dashboard, bedside display, and patient portal). We noted that the goal was not to compare PAM scores across services but to compare the scores before and after the intervention on the service level. To assess whether patient-reported characteristics could confound PAM score differences within each of the 3 services, we used a Fisher exact test to compare categorical variables across the 3 intervention groups and we used robust chi-square tests [30] (which do not assume normality) to compare continuous variables across the 3 groups (see Multimedia Appendix 2). Differences in patient characteristics among the 3 groups were controlled for using a weighted propensity score analysis [34]. Within each of the 3 services, the propensities of patients being in the 3 groups were estimated using a multinomial logistic regression model that included the variables in the table in Multimedia Appendix 2 as covariates. Each patient was weighted by the inverse probability of being in their observed group, with the goal of balancing observable characteristics among the 3 groups within service. After propensity weighting, the balance among the characteristics in the groups is also given in Multimedia Appendix 2. In addition, after propensity weighting, within the service, comparisons of the ordinal PAM scores among groups were performed using robust weighted linear regression [30]. All analyses were performed using SAS version 9.4 (SAS Institute).

## Results

### Patient Characteristics and Participation

Of the 18,075 patients on our study units, 12,737 (12,737/18,075, 70.47%) patients met the inclusion criteria and 2974 (2974/12,737, 23.35%) patients were asked to use the portal (Figure 2). Of the patients who were approached by study staff, 1755 (1755/2974, 59.01%) patients were enrolled in the patient portal. The most frequent reasons patients cited for declining were that they were not interested (56.52%), were leaving the hospital soon (14.68%), it involved too much technology (14.52%), or they felt too sick or too tired (4.59%). Patients who enrolled to use the portal tended to be younger, less sick, less likely to have Medicare as an insurer, and were more likely to be registered for the Partners Healthcare enterprise ambulatory portal (Patient Gateway; Table 1). A total of 80.49% of users were patients whereas 19.51% were care partners, only 1.60% created multiple accounts. Of all the users, approximately 37% preferred to use their own devices over the tablet computers provided by the study team.

**Figure 2.** Recruitment flow diagram.



### Use of the Patient Portal

A total of 1637 patients and care partners were enrolled and received initial teaching on the portal. Approximately 65% of users did not use the portal beyond the first day, 20.28 used the portal for 2 days, and 14.66% used the portal for 3 or more days (Table 2). Most users (95.42%) accessed the portal from 1 to 4 days. On average, users logged into the portal at least once a day for 1.80 days (range: 1-32 days) and logged in 42.69% of the days that they had access during their hospitalization. The Test Results page was the most frequently visited, followed by My Care Team and Medications pages (Figure 3).

**Table 2.** Use of patient portal.

| Usage measure | Portal users[a], n=1637 | Source of measure and source results | | |
|---|---|---|---|---|
| Participated in teaching after enrollment, n (%) | 1637 (100.00) | —[b] | — | — |
| Accessed portal (after initial teaching), n (%) | — | Grossman et al, 2017 [28], 10 (100) | Wilcox et al, 2016 [29], 20 (70) | Woollen et al, 2016 [7], 14 (86) |
| **Accessed portal for, n (%)** | — | — | — | — |
| 1 day only | 1065 (65.06) | — | — | — |
| 2 days only | 332 (20.28) | — | — | — |
| 3 or more days | 240 (14.66) | — | — | — |
| **Accessed portal for, n (%)** | — | | | |
| 1-4 days | 1562 (95.42) | Dalal et al, 2016 [27], 200 (84), n=239 | — | — |
| 5-10 days | 61 (3.73) | Dalal et al, 2016 [27], 39 (16), n=239 | — | — |
| >10 days | 14 (0.86) | — | — | — |
| **Days used (logged in at least once per day)** | — | — | — | — |
| Mean (SD) | 1.80 (2.28) | Runaas et al, 2018 [22], 7.6 (6.3), n=20 | — | — |
| Range | 1-32 | — | — | — |
| **Days users had access (during hospitalization)** | — | — | — | — |
| Mean (SD) | 6.20 (7.24) | Runaas et al, 2018 [22], 21.3[c], n-=20 | Grossman et al, 2017 [28], 13.3[c], n=20 | Masterson-Creber et al, 2018 [20][d], n=426 |
| Range | 1-66 | Runaas et al, 2018 [22], 15-37, n=20 | Grossman et al, 2017 [28], 4-38, n=20 | — |
| **Percentage of days used during access period (during hospitalization)** | — | — | — | — |
| Mean (SD) | 42.69 (27.71) | Dykes et al, 2017 [5], Grossman et al, 2018 [11], *Brigham and Women's Hospital patient-centered toolkit*,63[c], n=194 | — | — |
| Range | 1.5-100 | — | — | — |

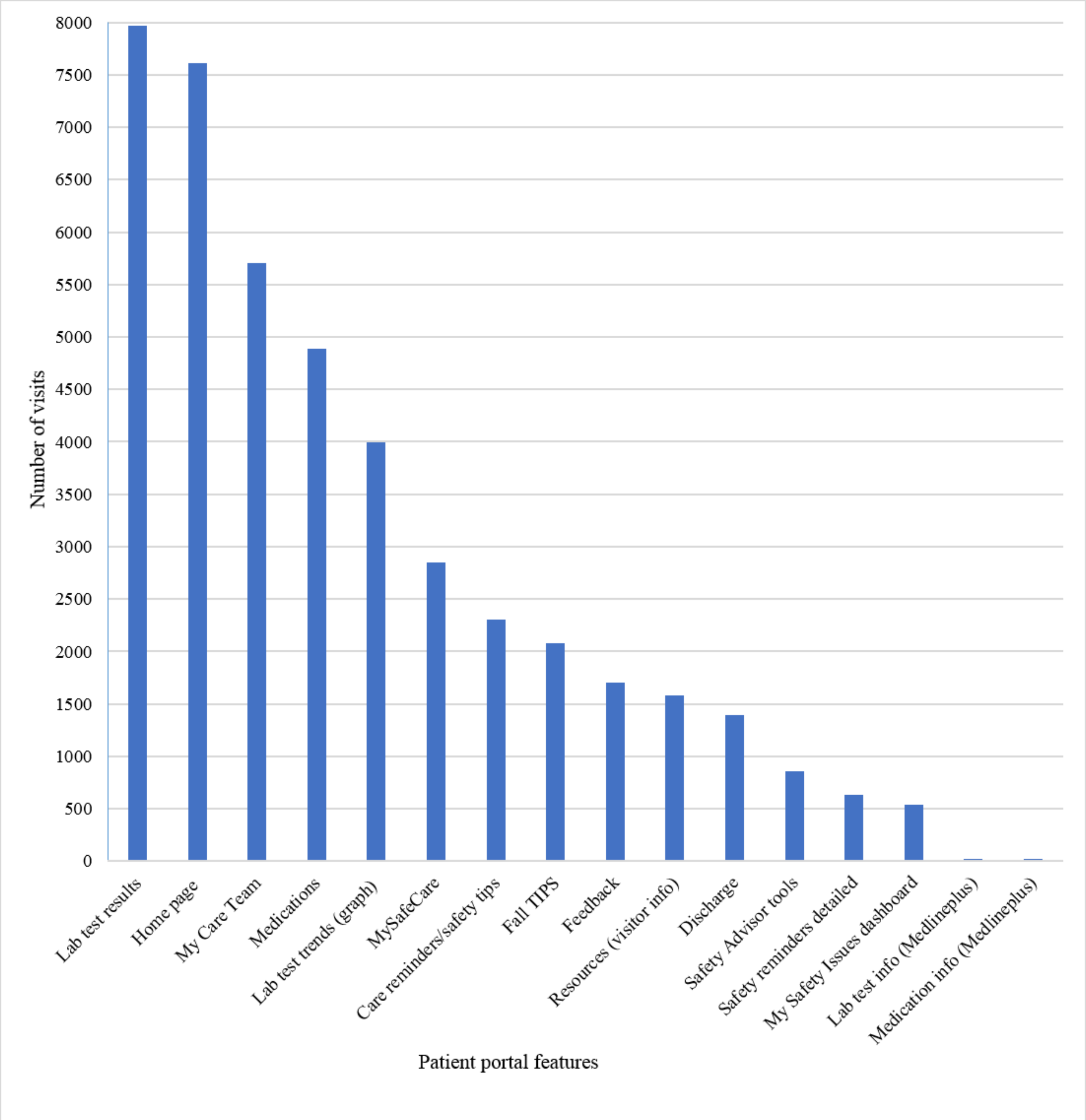
[a]Number of patient portal users—there can be more than one portal user per patient enrolled. Total users n=1637 (patients with 1 user only, n=1602; 2 users only, n=16; and 3 users, n=1).

[b]No data.

[c]SD is not available.

[e]median=3.17.

**Figure 3.** Patient portal use by feature. Some features were added after initial implementation, such as discharge was added on November 29, 2017, and the safety advisor (my safety issues dashboard) was added on January 1, 2018.




## Patient and Caregiver Activation

There was an increase in PAM scores between the preintervention (usual care) group and the postintervention (safety dashboard and bedside safety display only) group on the neurology and medicine services but not on the oncology service (Table 3). On the oncology service, the mean PAM score decreased in the safety dashboard and bedside safety display only group, but the PAM scores in the patient portal group increased when compared with the usual care group. On the medicine service, the PAM scores increased in the safety dashboard and bedside safety display only group; however, there was a nonsignificant decrease in the patient portal group. All services showed an increase in mean PAM score in the patient portal group compared with the usual care group. Overall, after propensity weighting, the increase in PAM scores was statistically significant on the neurology and general medicine services. Although the sample size was too small to make formal comparisons, the observed CG-PAM score trends were similar to those seen with the PAM scores.

**Table 3.** Patient activation measure survey outcomes (propensity weighted).

| Service | Usual care (preintervention[a]) | | Safety dashboard + bedside safety display only (postintervention[b]) | | Safety dashboard + bedside safety display + patient portal (postintervention) | | *P* value |
|---|---|---|---|---|---|---|---|
| | PAM[c]-13, n (CG-PAM[d]) | Mean (95% CI) | PAM-13, n (CG-PAM) | Mean (95% CI) | PAM-13, n (CG-PAM) | Mean (95% CI) | |
| Neurology | 124 (34) | 61.3 (58.7-63.9) | 127 (8) | 64.8 (61.8-67.8) | 251 (55) | 71.4 (69.2-73.6) | <.001 |
| Oncology | 122 (14) | 60.6 (57.6-63.0) | 33 (4) | 55.4 (45.9-64.9) | 71 (6) | 64.7 (60.3-69.2) | .14 |
| Medicine | 250 (21) | 61.8 (59.9-63.8) | 340 (8) | 66.1 (64.3-67.9) | 206 (6) | 65.5 (63.0-68.1) | .01 |

[a]Surveyed September to November 2016.

[b]Surveyed December 2016 to May 2018.

[c]PAM: patient activation measure.

[d]CG-PAM: caregiver patient activation measure.

## Discussion

### Principal Findings

We implemented a patient portal over a large set of inpatient units spanning different services and found that patient activation scores improved in association with access to the tools in addition to an increase in the group with access to an acute care patient portal. Approximately 60% of the patients and care partners approached chose to enroll in the patient portal. This is much higher than the 18% enrollment rate we saw in our previous study with the Patient Centered Toolkit (also from our organization, implemented on a medical intensive care unit and an oncology unit) [5]. Despite successfully enrolling more patients than a previous study [5], many patients still chose not to participate. The leading reasons were that they were simply not interested, whereas others felt overwhelmed by unfamiliar technology or felt too sick to participate. Similarly, our previous study of the Patient Centered Toolkit found that 2 frequent reasons given for declining were personal preference or pending discharge [27]. Although we did not formally investigate the reasons that patients were not interested in using the portal, anecdotally we observed that many patients cited the use of other ambulatory patient portals such as Patient Gateway and did not want to use a second portal. Interestingly, this did not match our comparison of patient characteristics between those that enrolled and those that declined, which found that enrolled patients were more likely to be Patient Gateway users. This potential barrier to adoption was noted previously, along with a lack of access to the portal outside the hospital for care partners, which we also experienced, and it was recommended that access to an acute care patient portal be offered through ambulatory portals [27]. Providing these options in the future may help engage more patients and care partners in using similar technology.

We leveraged existing patient portal usage measures to describe the use of the acute care patient portal among enrollees to conduct comparisons. Overall, our portal usage was not as high as seen in previous studies (Table 2). Many patients did not use the patient portal beyond the first day, although there was a wide range of time periods that patients accessed the portal. We found that patients used the portal for fewer days during the access period than we reported previously with the Patient Centered Toolkit [5]. We were not able to distinguish between use on the first day during and after the initial teaching, and therefore were not able to directly compare with other research describing spontaneous use after initial teaching [7,28,29]. The average duration of portal use was also lower than what Runaas et al reported in their evaluation of an acute care patient portal, but it is possible that this is a result of the longer lengths of stay of their bone marrow transplant patient population [22]. We did not study patients' perceptions of the portal or evaluate the usability of the portal during the intervention period. Patients may have found that the content was not useful to them or may have discovered usability or technical issues in accessing the portal; however, we met regularly with research staff who were enrolling patients on the clinical units and did not hear reports of usability issues from these staff.

We noted that the sample sizes for previous patient portal studies were much smaller than our study, and there was not enough data to conduct direct comparisons. Most of the previous studies were conducted as feasibility studies with small sample sizes. In contrast, our portal was implemented as a clinical trial, we approached close to 3000 patients and had limited resources for user support and follow-up. Although our study's portal use demonstrated lower usage, we suspect one of the contributing factors may have been less engagement with users to encourage them to use the portal throughout their hospital stay. Our usage data imply that a one-time engagement with patients during portal enrollment was not enough to encourage continued use of the portal throughout the patients' hospital stay. We have learned that ongoing support implemented into a clinical workflow may be key to successful implementation of an acute care patient portal and sustaining use beyond the research study.

Patients most often viewed their test results, care team members, and medication lists, similar to findings of other studies of acute care patient portals [27,28]. The tailored patient safety educational features that were unique to our patient-centered portal were not visited as often as we expected. We anticipated that these features could have impacted patient activation and safety. Although patients are interested in accessing their clinical information, we need a strategy to promote the use of additional portal features. Such strategies might include emphasizing the safety modules during teaching sessions or incorporating the use of the portal into formal patient education.

We found an association between patient activation and the use of patient-facing HIT tools in all study units, though not statistically significant in oncology. We hypothesized that the oncology service experienced different results because of organizational changes that occurred during our implementation period. In addition, many oncology patients were not able to participate in the PAM surveys and often had longer lengths of stay than patients on other services, contributing to the small sample size of oncology patients. This may have affected the results. The introduction of the bedside safety display on the neurology and medicine units may have led to the improved patient activation, and the use of an acute care patient portal in addition to the bedside safety display may have further improved patient activation. It is also possible that the patient groups surveyed had differing levels of activation at baseline; the patients who agreed to use the patient portal may have been more activated at baseline. This assumption was not always true based on our results—the bedside safety display group was more activated than the patient portal group in the medicine service. Overall, patient portals may help engage hospitalized patients and encourage active participation in care.

### Limitations

This study has several limitations: it was designed as a pre-post trial; it was conducted at a single academic medical center; the tools were independently designed; and it was only accessible to English-speaking patients or care partners. Additional studies are needed to assess the generalizability of our findings. We aimed to assess patient activation for patients using the acute patient portal; however, the other HIT tools such as the provider-facing safety dashboard and bedside patient display were implemented for all intervention groups and, therefore, we could not evaluate the effect of portal use on its own. We found an association between patient portal use and patient activation, but this may have been attributed to use and exposure to other components of the intervention.

In addition, our database had a technical issue causing some patient portal usage data to be lost; however, this was less than 10% of participants and other than length of stay, patient characteristics were similar to those of other portal users. Finally, our mobile app was only available on Apple smartphones, limiting some users that may have preferred to use their own mobile device.

### Conclusions

We found an association between the use of HIT tools, including a patient portal and patient safety display, and improved levels of patient activation in the inpatient setting. Such tools may be an effective mechanism to engage patients in their care and improve outcomes. Future study should continue to leverage existing usage metrics to assess patient portal use and focus on the impact of patient portals on specific outcome measures in the hospital setting.


## Acknowledgments

This study was supported by a grant from the AHRQ: 1P30HS0235335 Making Acute Care More Patient Centered. The funders had no role in the design or conduct of the study; collection, analysis, or interpretation of data; or preparation or review of the manuscript. The authors also thank all members of the BWH PSLL research team.


## Conflicts of Interest

DWB consults for EarlySense that makes patient safety monitoring systems. He receives cash compensation from CDI Negev Ltd, a not-for-profit incubator for HIT startups. He receives equity from Valera Health that makes software to help patients with chronic diseases. He receives equity from Clew that makes software to support clinical decision making in intensive care. He receives equity from MDClone that takes clinical data and produces deidentified versions of it. DWB's financial interests have been reviewed by BWH and Partners HealthCare in accordance with their institutional policies.

## Multimedia Appendix 1

Detailed description of patient portal features.

[PDF File (Adobe PDF File), 1MB - jmir_v21i7e13336_app1.pdf ]

## Multimedia Appendix 2

Demographics for patient activation measure survey (control factors).

[PDF File (Adobe PDF File), 140KB - jmir_v21i7e13336_app2.pdf ]

## References


1. Bickmore TW, Pfeifer LM, Jack BW. Taking the Time to Care: Empowering Low Health Literacy Hospital Patients With Virtual Nurse Agents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4-9, 2009; Boston, MA p. 1265-1274. [doi: 10.1145/1518701.1518891]
2. Mishra SR, Haldar S, Pollack AH, Kendall L, Miller AD, Khelifi M, et al. 'Not just a receiver': understanding patient behavior in the hospital environment. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May 7;2016:3103-3114 [FREE Full text] [doi: 10.1145/2858036.2858167] [Medline: 28345079]

3. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
4. Prey JE, Woollen J, Wilcox L, Sackeim AD, Hripcsak G, Bakken S, et al. Patient engagement in the inpatient setting: a systematic review. J Am Med Inform Assoc 2014;21(4):742-750 [FREE Full text] [doi: 10.1136/amiajnl-2013-002141] [Medline: 24272163]
5. Dykes PC, Rozenblum R, Dalal A, Massaro A, Chang F, Clements M, et al. Prospective evaluation of a multifaceted intervention to improve outcomes in intensive care: the promoting respect and ongoing safety through patient engagement communication and technology study. Crit Care Med 2017 Aug;45(8):e806-e813. [doi: 10.1097/CCM.0000000000002449] [Medline: 28471886]
6. Prey JE, Restaino S, Vawdrey DK. Providing hospital patients with access to their medical records. AMIA Annu Symp Proc 2014;2014:1884-1893 [FREE Full text] [Medline: 25954461]
7. Woollen J, Prey J, Wilcox L, Sackeim A, Restaino S, Raza ST, et al. Patient experiences using an inpatient personal health record. Appl Clin Inform 2016;7(2):446-460 [FREE Full text] [doi: 10.4338/ACI-2015-10-RA-0130] [Medline: 27437053]
8. Vawdrey DK, Wilcox LG, Collins SA, Bakken S, Feiner S, Boyer A, et al. A tablet computer application for patients to participate in their hospital care. AMIA Annu Symp Proc 2011;2011:1428-1435 [FREE Full text] [Medline: 22195206]
9. Dykes PC, Carroll DL, Hurley AC, Benoit A, Chang F, Pozzar R, et al. Building and testing a patient-centric electronic bedside communication center. J Gerontol Nurs 2013;39(1):15-19. [doi: 10.3928/00989134-20121204-03] [Medline: 23244060]
10. MGMA. 2018. MGMA Stat: Most Practices Offer a Patient Portal URL: https://www.mgma.com/news-insights/quality-patient-experience/mgma-stat-most-practices-offer-a-patient-portal [accessed 2019-01-09] [WebCite Cache ID 75IuPWe1Q]
11. Grossman LV, Choi SW, Collins S, Dykes PC, O'Leary KJ, Rizer M, et al. Implementation of acute care patient portals: recommendations on utility and use from six early adopters. J Am Med Inform Assoc 2018;25(4):370-379. [doi: 10.1093/jamia/ocx074] [Medline: 29040634]
12. O'Leary KJ, Sharma RK, Killarney A, O'Hara LS, Lohman ME, Culver E, et al. Patients' and healthcare providers' perceptions of a mobile portal application for hospitalized patients. BMC Med Inform Decis Mak 2016;16(1):123 [FREE Full text] [doi: 10.1186/s12911-016-0363-7] [Medline: 27653854]
13. Kelly MM, Hoonakker PL, Dean SM. Using an inpatient portal to engage families in pediatric hospital care. J Am Med Inform Assoc 2017;24(1):153-161. [doi: 10.1093/jamia/ocw070] [Medline: 27301746]
14. Kelly MM, Coller RJ, Hoonakker PL. Inpatient portals for hospitalized patients and caregivers: a systematic review. J Hosp Med 2018;13(6):405-412 [FREE Full text] [doi: 10.12788/jhm.2894] [Medline: 29261819]
15. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
16. Dentzer S. Rx for the 'blockbuster drug' of patient engagement. Health Aff (Millwood) 2013 Feb;32(2):202. [doi: 10.1377/hlthaff.2013.0037] [Medline: 23381509]
17. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
18. Mitchell SE, Gardiner PM, Sadikova E, Martin JM, Jack BW, Hibbard JH, et al. Patient activation and 30-day post-discharge hospital utilization. J Gen Intern Med 2014 Feb;29(2):349-355 [FREE Full text] [doi: 10.1007/s11606-013-2647-2] [Medline: 24091935]
19. Hibbard JH, Greene J, Overton V. Patients with lower activation associated with higher costs; delivery systems should know their patients' 'scores'. Health Aff (Millwood) 2013 Feb;32(2):216-222. [doi: 10.1377/hlthaff.2012.1064] [Medline: 23381513]
20. Masterson CR, Grossman LV, Ryan B, Qian M, Polubriaginof FC, Restaino S, et al. Engaging hospitalized patients with personalized health information: a randomized trial of an inpatient portal. J Am Med Inform Assoc 2019 Feb;26(2):115-123. [doi: 10.1093/jamia/ocy146] [Medline: 30534990]
21. O'Leary KJ, Lohman ME, Culver E, Killarney A, Randy SG, Liebovitz DM. The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. J Am Med Inform Assoc 2016;23(1):159-165. [doi: 10.1093/jamia/ocv058] [Medline: 26078412]
22. Runaas L, Hoodin F, Munaco A, Fauer A, Sankaran R, Churay T, et al. Novel health information technology tool use by adult patients undergoing allogeneic hematopoietic cell transplantation: longitudinal quantitative and qualitative patient-reported outcomes. JCO Clin Cancer Inform 2018 Dec;2:1-12 [FREE Full text] [doi: 10.1200/CCI.17.00110] [Medline: 30652535]
23. Mlaver E, Schnipper JL, Boxer RB, Breuer DJ, Gershanik EF, Dykes PC, et al. User-centered collaborative design and development of an inpatient safety dashboard. Jt Comm J Qual Patient Saf 2017;43(12):676-685. [doi: 10.1016/j.jcjq.2017.05.010] [Medline: 29173289]

24. Duckworth M, Leung E, Fuller T, Espares J, Couture B, Chang F, et al. Nurse, patient, and care partner perceptions of a personalized safety plan screensaver. J Gerontol Nurs 2017 Apr 1;43(4):15-22. [doi: 10.3928/00989134-20170313-05] [Medline: 28358972]
25. Dykes PC, Carroll DL, Hurley A, Lipsitz S, Benoit A, Chang F, et al. Fall prevention in acute care hospitals: a randomized trial. J Am Med Assoc 2010 Nov 3;304(17):1912-1918 [FREE Full text] [doi: 10.1001/jama.2010.1567] [Medline: 21045097]
26. Collins SA, Couture B, Smith AD, Gershanik EF, Lilley E, Chang F, et al. Mixed-methods evaluation of real-time safety reporting by hospitalized patients and their care partners: the MySafeCare application. J Patient Saf 2018 Apr 27:-(forthcoming). [doi: 10.1097/PTS.0000000000000493] [Medline: 29781978]
27. Dalal AK, Dykes PC, Collins S, Lehmann LS, Ohashi K, Rozenblum R, et al. A web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: a preliminary evaluation. J Am Med Inform Assoc 2016;23(1):80-87. [doi: 10.1093/jamia/ocv093] [Medline: 26239859]
28. Grossman LV, Creber RM, Restaino S, Vawdrey DK. Sharing clinical notes with hospitalized patients via an acute care portal. AMIA Annu Symp Proc 2017;2017:800-809 [FREE Full text] [Medline: 29854146]
29. Wilcox L, Woollen J, Prey J, Restaino S, Bakken S, Feiner S, et al. Interactive tools for inpatient medication tracking: a multi-phase study with cardiothoracic surgery patients. J Am Med Inform Assoc 2016;23(1):144-158 [FREE Full text] [doi: 10.1093/jamia/ocv160] [Medline: 26744489]
30. Lipsitz SR. Methods for estimating the parameters of a linear model for ordered categorical data. Biometrics 1992 Mar;48(1):271-281. [doi: 10.2307/2532755] [Medline: 1581487]
31. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
32. Solihull Together for Better Lives. 2015. Insignia Health: Caregiver/Carer Patient Activation Measure® (CG-PAM®) 13 UK URL: http://solihulltogether.co.uk/images/Patient_Activation/UK_Carer_PAM_13.pdf [accessed 2019-04-24] [WebCite Cache ID 77sWfHvjd]
33. Prey JE, Qian M, Restaino S, Hibbard J, Bakken S, Schnall R, et al. Reliability and validity of the patient activation measure in hospitalized patients. Patient Educ Couns 2016;99(12):2026-2033 [FREE Full text] [doi: 10.1016/j.pec.2016.06.029] [Medline: 27422339]
34. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997 Oct 15;127(8 Pt 2):757-763. [doi: 10.7326/0003-4819-127-8_Part_2-199710151-00064] [Medline: 9382394]

## Abbreviations

**AHRQ:** Agency for Healthcare Research and Quality
**BWH:** Brigham and Women's Hospital
**CG-PAM:** caregiver patient activation measure
**EHR:** electronic health record
**HIT:** health information technology
**PAM:** patient activation measure
**PSLL:** Patient Safety Learning Laboratory

*Edited by G Eysenbach; submitted 10.01.19; peer-reviewed by M Kelly, J Hibbard; comments to author 14.03.19; revised version received 08.05.19; accepted 10.05.19; published 18.07.19.*

*Please cite as:*
*Schnock KO, Snyder JE, Fuller TE, Duckworth M, Grant M, Yoon C, Lipsitz S, Dalal AK, Bates DW, Dykes PC*
*Acute Care Patient Portal Intervention: Portal Use and Patient Activation*
*J Med Internet Res 2019;21(7):e13336*
*URL: https://www.jmir.org/2019/7/e13336/*
*doi:10.2196/13336*
*PMID:31322123*

Original Paper

# Validation of the Electronic Version of the International Index of Erectile Function (IIEF-5 and IIEF-15): A Crossover Study

Rob A A van Kollenburg[1], MD; Daniel Martijn de Bruin[1], PhD; Hessel Wijkstra[1,2], PhD, IR

[1]Department of Urology, Amsterdam University Medical Centers, Location Amsterdam Medical Center, University of Amsterdam, Amsterdam, Netherlands

[2]Signal Processing Systems, Eindhoven University of Technology, Eindhoven, Netherlands

**Corresponding Author:**
Rob A A van Kollenburg, MD
Department of Urology
Amsterdam University Medical Centers, Location Amsterdam Medical Center
University of Amsterdam
Meibergdreef 9
Amsterdam, 1105 AZ
Netherlands
Phone: 31 205666465
Email: r.a.vankollenburg@amsterdamumc.nl

## Abstract

**Background:** Patient-reported outcome measures (PROMs) are increasingly used to measure patient's perspective of functional well-being, disease burden, treatment effectiveness, and clinical decision making. Electronic versions are increasingly feasible because of smartphone and tablet usage. However, validation of these electronic PROMs (ePROMs) is warranted for justified implementation. The International Index of Erectile Function (IIEF) 5 and 15 are widely used PROMs in urology to measure erectile dysfunction. Measurement reliability and validity testing of the IIEF ePROMs are essential before clinical application.

**Objective:** The aim of this study was to assess reliability and validity of an ePROM version of both IIEF-5 and 15.

**Methods:** This study included 179 patients from our urology outpatient clinic. It also had a randomized crossover design—participants completed either a paper and electronic IIEF-5 or 15 or twice completed an electronic version—with a 5-day delay. Internal consistency was assessed using Cronbach alpha and Spearman-Brown coefficient, test-retest reliability using the intraclass correlation coefficient (ICC), and convergent validity using the Pearson and Spearman correlation coefficient.

**Results:** A total of 122 participants completed the study. Internal consistency was excellent for the electronic IIEF-5 (ICC 0.902) and good to excellent for the domains of the IIEF-15 (ICC 0.962-0.834). Test-retest reliability was excellent for the IIEF-5 (ICC 0.924) and good to excellent for the domains of the IIEF-15 (ICC 0.950-0.778). Convergent validity was excellent for the IIEF-5 and IIEF-15, with a correlation of r=0.923 and r=0.951, respectively.

**Conclusions:** We successfully introduced patient-acceptable ePROM versions of the IIEF-5 and IIEF-15. This study's results demonstrate that the ePROM versions of the IIEF-5 and IIEF-15 can be reliably implemented, as outcomes are reliable and in accordance with findings of the paper version.

**Trial Registration:** ClinicalTrials.gov NCT03222388; https://clinicaltrials.gov/ct2/show/NCT03222388

***(J Med Internet Res 2019;21(7):e13490)*** doi:10.2196/13490

**KEYWORDS**

ePROM; smartphone; surveys and questionnaires

## Introduction

### Background

The International Index for Erectile Function (IIEF) is a patient-reported outcome measure (PROM), widely used in urology to measure erectile dysfunction (ED), applied both in clinical research and in daily clinical practice [1]. The 15-item version was developed by Rosen in 1997, and a 5-item short version followed in 1999 [2,3]. Translations into over 32 languages and validation of these translations followed [1,4,5]. Electronic PROMs (ePROMs), the electronic version of PROMs, are increasingly used, as the internet is easily accessible through mobile devices. The standard PROM is shifting from conventional paper and pen toward electronic administration, making ePROMs the (upcoming) new standard [6]. Attributing factors are smartphone use and subsequent development of patient-focused apps. Advantages of electronic administration are feasibility, automated calculations, reduced missing and ambiguous data, and increased compliance [7]. However, simple digitalization of existing PROMs does not assure reliability of ePROMs as administration, and subsequently outcomes, may be altered [6,8]. Therefore, reliability testing is advised to assure quality of ePROMs [8]. The extent of ePROM testing depends on the changes made during the PROM to ePROM transformation. Layout changes, for example, splitting the format into single questions, can be classified as a moderate level of modification [8]. For moderate-level modifications, a formal equivalence assessment of the electronic measure is advised, to show no significant difference in paper and electronic PROM scoring [8]. Given the fact that smartphone- and tablet-feasible ePROM versions of the IIEF-5 and 15 will probably include layout changes, reliability and validity testing of the IIEFs is therefore needed to assure outcome quality.

### Aim

The primary objective of this study was to develop an ePROM version of both the IIEF-5 and IIEF-15 and test reliability and validity in a male population.

## Methods

This observational study was conducted in a tertiary medical center, the Amsterdam University Medical Centers (UMCs), location Amsterdam Medical Center. The study received an ethics review waiver from the Institutional Review Board (W17.281), and the study was registered on Clinical Trial.gov (NCT03222388).

### Study Population

Male patients visiting the outpatient clinic of the urologic department were eligible for participation, patients were enrolled during a 6-month period, from July 2017 to December 2017. Screening for study eligibility (eg, inclusion and exclusion criteria and general health status) was based on information in the electronic patient file. Screening was performed by a medical doctor (RK, the primary author). Eligible patients were approached at the outpatient clinic before consultation. When interested, patients were informed about the study, and written informed consent was obtained. Inclusion criteria comprised males ≥40 years of age, in possession of an electronic device (smartphone/tablet/laptop), and fluent in Dutch. Exclusion criteria were adjustment of treatment during consultation (especially ED treatment), unable to provide informed consent, or unfit according to the medical doctor (eg, poor general health status).

### International Index for Erectile Function 5 and 15

The IIEF-15 comprises 15 items divided into 5 domains: erectile function, orgasmic function, sexual desire, intercourse satisfaction, and overall satisfaction, respectively. The IIEF-5 comprises 5 items from the IIEF-15, 4 from the erectile function domain, and 1 from intercourse satisfaction. Response options for each item ranged from 1 to 5, and occasionally the option "0," depicting no sexual stimulation/intercourse. Scores are summed. Both versions have official Dutch translations [2,4].

### Study Design

A total of 179 participants were randomly assigned by the database management system (DMS) to the IIEF-5 or IIEF-15. Participants were hereafter randomly assigned to 2 groups: electronic version followed by electronic version (EE) or paper version and electronic version (PE). Primarily, participants in the PE group would randomly fill out either the paper or electronic IIEF to correct for order effects. This resulted in 6 different groups: (1) IIEF-5 paper electronic, (2) IIEF-5 electronic paper, (3) IIEF-5 electronic, (4) IIEF-15 paper electronic, (5) IIEF-15 electronic paper, and (6) IIEF-15 electronic. Participants were stratified on the basis of age <60 or ≥60, to improve group homogeneity on the basis of expected experience with internet/mobile devices.

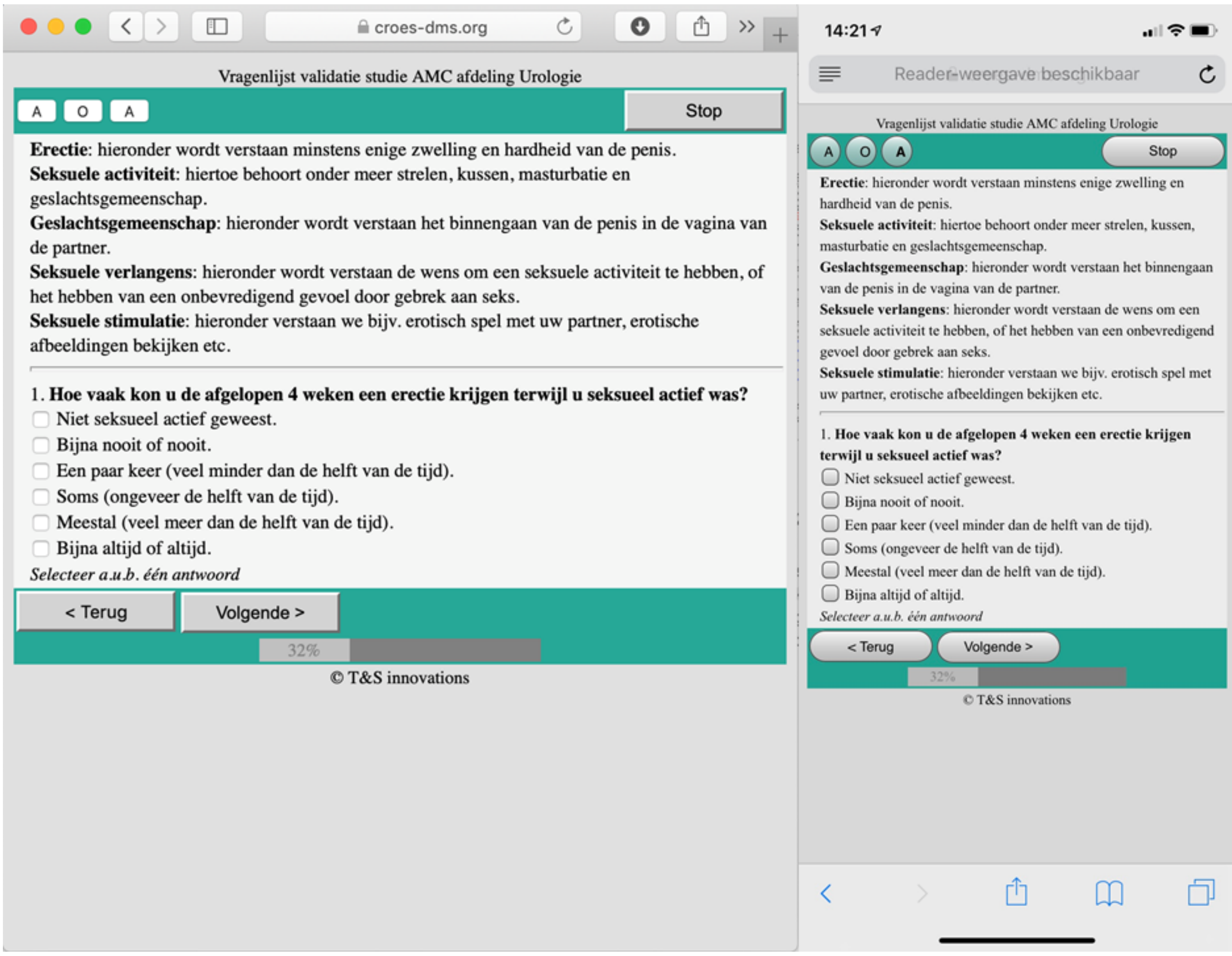
### Study Methods

Participants assigned to a group with paper IIEF received this PROM in a sealed envelope during inclusion. The paper IIEF was returned to the researcher by an included return envelope with a stamp. Received paper IIEFs were coded and data were entered in the DMS. In case of missing data, the input was left blank. Participants received 2 emails containing a link to the ePROM, which could be completed at home at any convenient moment. The first invitation was sent 1 day postinclusion. A second invitation was sent 5 days after completion of the first ePROM. Reminders were sent twice, with a 3-day delay. If necessary, a personal reminder followed. The emails contained a link that redirected to a Web-based questionnaire. The first questionnaire started with several general questions, followed by either instructions for paper IIEF administration or the ePROM IIEF. This second questionnaire started with instruction or ePROM, followed by several evaluation questions.

### Electronic Patient-Reported Outcome Measure System

The electronic questionnaire system for IIEF administration was built as part of the DMS (available for specific users at ts-innovations.com) [9]. The system was equipped with an ePROM module and automated invitations. The system worked as a Web-based environment with an identical interface across platforms (eg, Safari, Chrome, computer, and smartphone). The system displayed one PROM item at a time, and the patient had to click for the next question. This made it possible to display almost all information on the screen, without the need for scrolling. A system preview is presented in Figure 1.

**Figure 1.** Screenshots of the ePROM displayed in a browser (Safari, left) and on a mobile device (iPhone, right).




## Electronic Patient-Reported Outcome Measure User Experience and Feasibility

After completion of the study, participants were asked about their willingness and preference to complete either only the PROM or ePROM or both. In addition, participants were asked to rate the overall ePROM quality on a scale of 1 to 10.

## Statistical Methods

Descriptive analyses were used for comparison of patient characteristics and feasibility outcomes. A 2-sided alpha level of .05 was considered statistically significant. Statistical analyses were performed using SPSS version 24.0 (SPSS inc).

## Sample Size

A total sample size of 172 participants was calculated for this study.

### *Sample Size: Paper Version and Electronic Version Groups*

A 2-sided 95% CI was computed using the large sample normal approximation for an intraclass correlation on the basis of 2 PROMs, and it will extend about 0.100 from the observed intraclass correlation when the expected intraclass correlation is 0.800. This resulted in a sample size of 51. Anticipating a 20% dropout resulted in a sample size of 61 participants per PROM, thus resulting in 122 participants in total.

### *Sample Size: Electronic Version Followed by Electronic Version*

A 2-sided 95% CI was computed using the large sample normal approximation for an intraclass correlation based on 2 PROMs, and it will extend about 0.100 from the observed intraclass correlation when the expected intraclass correlation is 0.880. This resulted in a sample size of 21. Anticipating a 20% dropout resulted in a sample size of 25 participants, thus resulting in 50 participants in total. The expected ICC of .88 was extracted from the Dutch IIEF-5 translation [4]. All sample sizes were calculated with the nQuery advisor software, provided by the Amsterdam UMC.

## Measurement Properties

The measurement properties were tested by the following methods:

1. The *internal consistency* is a measure of the extent to which items in a questionnaire scales and subscales are correlated, thus measuring the same concept [10]. The internal consistency was calculated for both paper and electronic IIEF data from the PE groups by Cronbach alpha or Spearman-Brown coefficient for 2-item subscales. An alpha ≥.9 reflected an excellent internal consistency, .9> alpha ≥ .8 reflected good consistency, and .8> alpha ≥.7 reflected acceptable internal consistency.

2. The *test-retest reliability* is the reliability of a test over time. The agreement between 2 repeated measurements was addressed with use of the ICC. These results were calculated based on the EE group results.
3. *Convergent validity* was also assessed. Support for this type of validity is provided if the total scale score and the subscale scores of the electronic version correlate substantially with the concerning scores of the original paper version. Convergent validity was analyzed using the Pearson correlation coefficient (*r*) or, when appropriate, Spearman rank correlation coefficient (rs) to determine the strength of the association between the paper and electronic IIEF.

For the ICC, a 2-way mixed-effect model, single measurement, and absolute agreement model was used. An ICC ≥0.9 reflected an excellent reliability, 0.9> ICC ≥0.75 reflected good reliability, and 0.75> ICC ≥0.5 reflected acceptable reliability, and <0.5 reflected poor reliability [11]. Pearson values *r* ≥0.5 reflected strong correlation, 0.5> *r*>0.3 reflected moderate correlation, and a 0.3> *r*>0.1 reflected weak correlation. A rank correlation of rs ≥0.5 reflected strong correlation, 0.5> rs >0.3 reflected moderate correlation, and 0.3> rs >0.1 reflected weak correlation [12].

### Data Safety

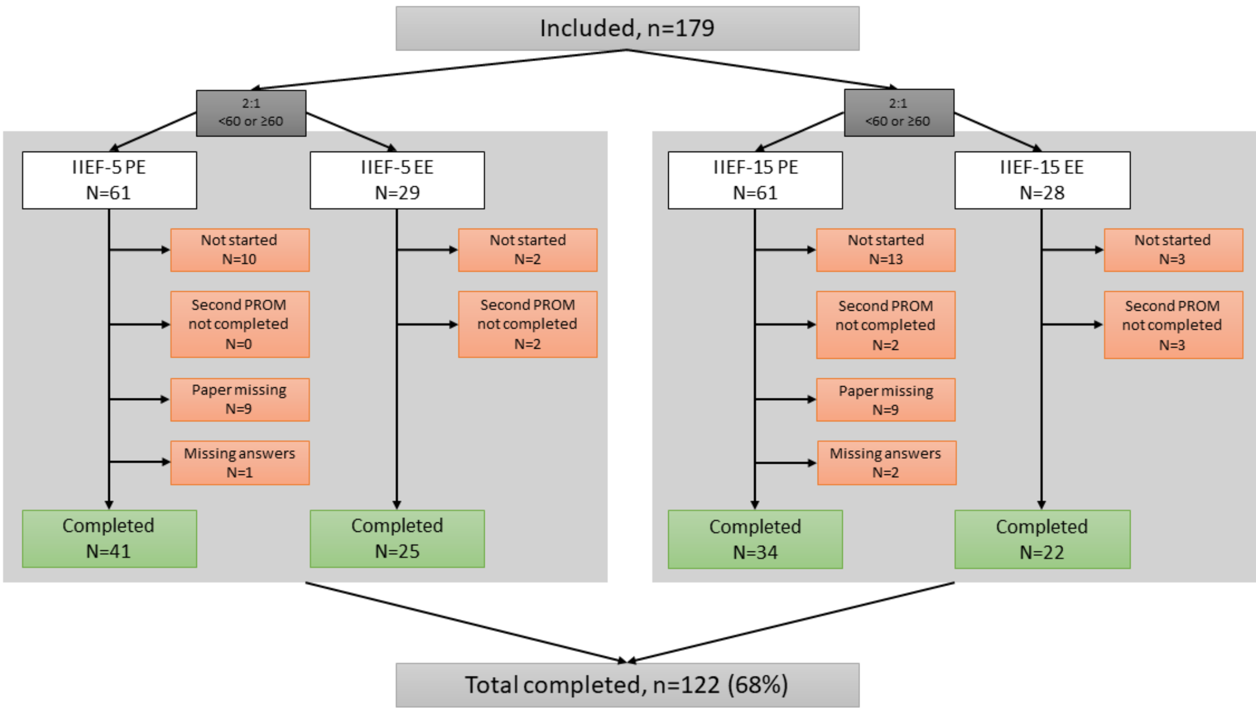
Data safety was guaranteed, as the emailed link redirected participants to a safe, validated, secured, Web-based environment. Information was directly stored in the DMS. No information was saved on the device itself, and all communication with the DMS was via an encrypted connection. The DMS was certified to store medical data (ISO9001, 14001, 27001:2013, and NEN7510). This was in line with Dutch guidelines and law concerning electronic collection of medical information.

## Results

### Participant Characteristics

A total of 179 men were included in this study. A total of 122 participants completed the study and were included in the final analysis. Figure 2 provides an overview of participant allocation over groups, number of participants who completed the study, and numbers and reasons for participant exclusion. The overall mean age was 61.3±9.5 years (range 41-81 years). An extensive overview of participant characteristics is available in Multimedia Appendix 1. The mean time between (e)PROM completion was 7.26±4.23 days, ranging from 5 to 32.

**Figure 2.** Participant inclusion and group allocation criteria (dark grey boxes). In the white boxes, the boxes contain the number of included participants per group. In the red boxes, the boxes contain excluded participants with reasons for exclusion. The green boxes show the number of included participants in the final analyses. EE: electronic version followed by electronic version; IIEF: International Index of Erectile Function; PE: paper version and electronic version; PROM: patient-reported outcome measure.




### Internal Consistency

The internal consistency of the IIEF-5 is excellent for both the paper and electronic version (Table 1). The internal consistency for the paper IIEF-15 domains is good to excellent, ranging from 0.846 to 0.971.

**Table 1.** Internal consistency measured by Cronbach alpha or Spearman-Brown coefficient.

| Measure | Paper | Electronic |
|---|---|---|
| IIEF[a]-5 | .954[b] | .902[b] |
| **IIEF-15** | **.974[b]** | **.840[b]** |
| Erectile function | .955[b] | .962[b] |
| Orgasmic function | 0.971[c] | 0.937[c] |
| Sexual desire | 0.887[c] | 0.848[c] |
| Intercourse satisfaction | .935[b] | .917[b] |
| Overall satisfaction | 0.890[c] | 0.924[c] |

[a]IIEF: International Index for Erectile Function.

[b]For Cronbach alpha.

[c]For Spearman-Brown coefficient.

**Table 2.** Reliability of the electronic International Index for Erectile Function, calculated with the intraclass correlation coefficient.

| Measure | Intraclass coefficient (95% CI) | *P* value |
|---|---|---|
| IIEF[a]-5 EE [b] (n=25) | 0.924 (0.837-0.966) | <.001 |
| **IIEF-15 EE (n=22)** | | |
| Erectile function | 0.933 (0.847-0.971) | <.001 |
| Orgasmic function | 0.778 (0.501-0.905) | <.001 |
| Sexual desire | 0.823 (0.619-0.923) | <.001 |
| Intercourse satisfaction | 0.950 (0.883-0.979) | <.001 |
| Overall satisfaction | 0.878 (0.733-0.947) | <.001 |

[a]IIEF: International Index for Erectile Function**.**

[b]EE: electronic version followed by electronic version.

## Test-Retest Reliability Electronic International Index for Erectile Function

The test-retest reliability of the electronic version of the IIEF-5 was excellent with an ICC of 0.924 and 95% CI of 0.837-0.966 (Table 2). For the IIEF-15, the test-retest reliability was excellent for the domains erectile function and intercourse satisfaction, with an ICC of 0.933 and 0.950, respectively. The domains orgasmic function, sexual desire, and overall satisfaction were good with an ICC of 0.778, 0.823, and 0.878, respectively. All calculated correlation coefficients were significant (*P*<.001).

## Convergent Validity

The convergent validity for the IIEF-5 calculated by Pearson correlation coefficient was *r*=0.923 (Table 3). The overall correlation for the IIEF-15 scale was excellent, *r*=0.951. The correlations for the IIEF-15 subdomains ranged from 0.987 to 0.900. All calculated correlations were excellent and significant (*P*<.001).

**Table 3.** Concurrent validity across the paper and electronic International Index for Erectile Function, calculated with the Pearson correlation coefficient and Spearman rank correlation coefficient.

| Measure | Correlation | *P* value |
|---|---|---|
| IIEF[a]-5 PE [b] (n=41) | 0.923[c] | <.001 |
| **IIEF-15 PE (n=34)** | **0.951[c]** | **<.001** |
| Erectile function | 0.987[c] | <.001 |
| Orgasmic function | 0.947[d] | <.001 |
| Sexual desire | 0.900[d] | <.001 |
| Intercourse satisfaction | 0.973[c] | <.001 |
| Overall satisfaction | 0.917[d] | <.001 |

[a]IIEF: International Index for Erectile Function.

[b]PE: paper version and electronic version.

[c]For Pearson correlation coefficient.

[d]For Spearman rank correlation coefficient.

**Table 4.** Feasibility outcomes.

| Evaluating question | IIEF[a]-5 | IIEF-15 | *P* value |
|---|---|---|---|
| Willingness to complete either paper, electronic, or both IIEF | Only electronic 6 (15%); Only paper 6 (15%); Both 27 (69%) | Only electronic 8 (26%); Only paper 1 (3%); Both 22 (71%) | .81 |
| Preference to complete either the paper or electronic IIEF | Electronic 25 (64%); Paper 8 (21%); None 6 (15%) | Electronic 18 (58%); Paper 8 (26%); None 5 (16%) | .52 |
| Electronic IIEF: overall rating | 7.8 (SD 1.3; range 4-10) | 7.8 (SD 1.0; range 6-10) | —[b] |

[a]IIEF: International Index of Erectile Function.

[b]Not applicable.

### Feasibility

Participants preferred an electronic version of the IIEF. After completion of both the PROM and ePROM IIEF, 69% of the IIEF-5 and 71% of the IIEF-15 participants were willing to complete both paper and electronic versions (Table 4). A vast majority preferred of the electronic versions with 64% and 58%, respectively. These numbers are similar to other studies [8]. Overall rating was 7.8 for both the IIEF-5 and IIEF-15.

### Participant Dropout

The actual number of participant dropout was higher than expected during sample size calculation. The actual number is 57 (32%), compared with the expected number of 28 (20%). A considerable number of participant dropout was a consequence of participants not starting at all (n=28, 49%) and paper IIEF's not received by the authors (n=19, 33%). All reasons for dropout and missing data are shown in Figure 2.

## Discussion

### Principal Findings

The objective of this study was to develop ePROM versions of the IIEF-5 and 15 and test reliability and validity. The findings from this study demonstrated that both the electronic IIEF-5 and the IIEF-15 showed good-to-excellent internal consistency, test-retest reliability, and convergent validity to their paper version.

### Comparison With Literature

Outcomes of this study are in line with outcomes of previous validation studies of related PROMs. Reliability outcomes are in accordance with literature. The ICC of 0.924 for the IIEF-5 is in line with the ICC of 0.960 found in earlier research on electronic testing [13]. The ICC outcomes for the IIEF-15 ranging from 0.950-0.778 are in line with expectations of descriptive literature [8]. Findings are also in line with other review articles that compared ePROM validation outcomes [6,7]. It can be argued that the electronic IIEF-5 validation was redundant, as it was already shown on personal digital assistant (PDA) by Matthew et al [13]. However, a smartphone/computer differs from a PDA interface, and the study of Matthew at al used an interval of 30 min, whereas a washout period of at least 2 days is advised [8]. Therefore, we decided to include the IIEF-5 as well. Feasibility outcomes show that participants were willing to fill both versions, with a preference for the electronic version. This is in line with the increasing interest for ePROMs and their validation.

### Strengths and Limitations

The strengths of this study are the time between administrations, inclusion of the test-retest group, and administration at home. As other studies complied with a time delay of 30 min between administration moments, this study had a 5-day delay [6]. This reduced carryover effects, and this thus improved quality and reliability of the outcomes [8]. Furthermore, we decided to include a group that administrated the questionnaire twice electronically, hereby we could show the test-retest reliability of the electronic versions. A last strength of this study is the moment of administration. As invitations were sent via email, participants could complete the IIEF at home. This resulted in a standardized administration environment, which is identical to future administration factors; this improved the data quality. The limitations of this study concern the included population and dropout numbers. For this study, we chose the general population of our outpatient clinic. This resulted in a more heterogeneous population than specifically men with consultation for possible ED, the intended IIEF population. Men who are not sexually active were also included. We reasoned that this would not be a problem, as the objective was to show reliability and validity of the electronic IIEF version. Other issues that need to be addressed are the dropout numbers. The actual dropout number was higher (n=57, 32%) than anticipated (20%). All factors are shown in Figure 2. However, it is reasonable to assume that the missing data would not have significantly impacted the study outcome, as the obtained results were significant and in line with literature. The outcomes of this study are useful, as ePROMs are becoming more important in daily practice. For urology, it is likely that the ePROM version of the IIEFs will be used in clinical and research settings in the near future. Outcomes of this study are representable for IIEF application as ePROM as long as item presentation is in a similar, sequential manner.

## Conclusions

This study, with a randomized crossover design, demonstrated that the electronic IIEF-5 and IIEF-15 showed equivalence to the paper version. Electronic versions can therefore be used reliably in clinical and research settings. Outcomes are reliable and in accordance with findings of the paper version.


## Acknowledgments

This work was supported by a grant from Cure for Cancer.


## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Complete overview of all participant characteristics.

[DOCX File, 54KB - jmir_v21i6e13490_app1.docx ]

## Multimedia Appendix 2

All individual items of the paper-electronic and test-retest groups of the IIEF-5 and IIEF-15. IIEF: International Index of Erectile Function.

[PDF File (Adobe PDF File), 40KB - jmir_v21i7e13490_app2.pdf ]

## References


1. Narang GL, Pannell SC, Laviana AA, Huen KH, Izard J, Smith AB, et al. Patient-reported outcome measures in urology. Curr Opin Urol 2017 Jul;27(4):366-374. [doi: 10.1097/MOU.0000000000000412] [Medline: 28505039]
2. Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. Urology 1997 Jun;49(6):822-830. [Medline: 9187685]
3. Rosen RC, Cappelleri JC, Smith MD, Lipsky J, Peña BM. Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. Int J Impot Res 1999 Dec;11(6):319-326. [Medline: 10637462]
4. Utomo E, Blok BF, Pastoor H, Bangma CH, Korfage IJ. The measurement properties of the five-item International Index of Erectile Function (IIEF-5): a Dutch validation study. Andrology 2015 Nov;3(6):1154-1159 [FREE Full text] [doi: 10.1111/andr.12112] [Medline: 26453539]
5. Rosen RC, Cappelleri JC, Gendrano N. The International Index of Erectile Function (IIEF): a state-of-the-science review. Int J Impot Res 2002 Aug;14(4):226-244 [FREE Full text] [doi: 10.1038/sj.ijir.3900857] [Medline: 12152111]
6. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015 Oct 07;13:167 [FREE Full text] [doi: 10.1186/s12955-015-0362-x] [Medline: 26446159]
7. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333. [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]

8. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009 Jun;12(4):419-429. [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
9. Wijkstra H. And its name is? DMS Version 3.0.x. J Endourol 2016 Dec;30(5):489-492. [doi: 10.1089/end.2016.29013.cro] [Medline: 27152909]
10. 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]
11. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
12. Cohen J. Statistical Power Analysis For The Behavioral Sciences (2nd Edition). London: Routledge; 1988.
13. Matthew AG, Currie KL, Irvine J, Ritvo P, Santa MD, Jamnicky L, et al. Serial personal digital assistant data capture of health-related quality of life: a randomized controlled trial in a prostate cancer clinic. Health Qual Life Outcomes 2007 Jul 06;5:38 [FREE Full text] [doi: 10.1186/1477-7525-5-38] [Medline: 17617906]


## Abbreviations

**DMS:** database management system
**ED:** erectile dysfunction
**EE:** electronic version followed by electronic version
**ePROM:** electronic patient-reported outcome measure
**IIEF:** International Index of Erectile Function
**PDA:** personal digital assistant
**PE:** paper version and electronic version
**PROM:** patient-reported outcome measure
**UMC:** University Medical Center


*Edited by G Eysenbach; submitted 25.01.19; peer-reviewed by N Azevedo, M Leveridge, JR Bautista; comments to author 28.02.19; revised version received 25.04.19; accepted 07.05.19; published 02.07.19.*

*Please cite as:*
*van Kollenburg RAA, de Bruin DM, Wijkstra H*
*Validation of the Electronic Version of the International Index of Erectile Function (IIEF-5 and IIEF-15): A Crossover Study*
*J Med Internet Res 2019;21(7):e13490*
*URL: https://www.jmir.org/2019/7/e13490/*
*doi:10.2196/13490*
*PMID:31267983*

Original Paper

# Development of In-Browser Simulators for Medical Education: Introduction of a Novel Software Toolchain

Jan Šilar[1,2], Ing; David Polák[1,2], Bc; Arnošt Mládek[1,2], RNDr, PhD; Filip Ježek[1], Ing; Theodore W Kurtz[3], MD; Stephen E DiCarlo[4], PhD; Jan Živný[1], Doc, MUDr, PhD; Jiri Kofranek[1,2], Doc, MUDr, CSc

[1]Institute of Pathological Physiology, First Faculty of Medicine, Charles University, Prague, Czech Republic

[2]Creative Connections s.r.o., Prague, Czech Republic

[3]Department of Laboratory Medicine, University of California, San Francisco, CA, United States

[4]Department of Physiology, College of Osteopathic Medicine, Michigan State University, East Lansing, MI, United States

**Corresponding Author:**
Jiri Kofranek, Doc, MUDr, CSc
Institute of Pathological Physiology
First Faculty of Medicine
Charles University
U Nemocnice 5
Praha 2
Prague, 128-53
Czech Republic
Phone: 420 777686868
Email: kofranek@gmail.com

## Abstract

**Background:** Simulators used in teaching are interactive applications comprising a mathematical model of the system under study and a graphical user interface (GUI) that allows the user to control the model inputs and visualize the model results in an intuitive and educational way. Well-designed simulators promote active learning, enhance problem-solving skills, and encourage collaboration and small group discussion. However, creating simulators for teaching purposes is a challenging process that requires many contributors including educators, modelers, graphic designers, and programmers. The availability of a toolchain of user-friendly software tools for building simulators can facilitate this complex task.

**Objective:** This paper aimed to describe an open-source software toolchain termed Bodylight.js that facilitates the creation of browser-based client-side simulators for teaching purposes, which are platform independent, do not require any installation, and can work offline. The toolchain interconnects state-of-the-art modeling tools with current Web technologies and is designed to be resilient to future changes in the software ecosystem.

**Methods:** We used several open-source Web technologies, namely, WebAssembly and JavaScript, combined with the power of the Modelica modeling language and deployed them on the internet with interactive animations built using Adobe Animate.

**Results:** Models are implemented in the Modelica language using either OpenModelica or Dassault Systèmes Dymola and exported to a standardized Functional Mock-up Unit (FMU) to ensure future compatibility. The C code from the FMU is further compiled to WebAssembly using Emscripten. Industry-standard Adobe Animate is used to create interactive animations. A new tool called Bodylight.js Composer was developed for the toolchain that enables one to create the final simulator by composing the GUI using animations, plots, and control elements in a drag-and-drop style and binding them to the model variables. The resulting simulators are stand-alone HyperText Markup Language files including JavaScript and WebAssembly. Several simulators for physiology education were created using the Bodylight.js toolchain and have been received with general acclaim by teachers and students alike, thus validating our approach. The Nephron, Circulation, and Pressure-Volume Loop simulators are presented in this paper. Bodylight.js is licensed under General Public License 3.0 and is free for anyone to use.

**Conclusions:** Bodylight.js enables us to effectively develop teaching simulators. Armed with this technology, we intend to focus on the development of new simulators and interactive textbooks for medical education. Bodylight.js usage is not limited to developing simulators for medical education and can facilitate the development of simulators for teaching complex topics in a variety of different fields.

*(J Med Internet Res 2019;21(7):e14160)* doi:10.2196/14160

**KEYWORDS**

education; physiology; computer simulation; modeling; Web browser; Web technologies

## Introduction

### Background

Educators are tasked to develop innovative and creative educational materials that supplement and further enhance the traditional lecture format. This requires developing and disseminating materials that facilitate active learning, enhance problem-solving skills, and encourage discussion and interaction in small group environments. Computer simulations are one way to fulfill these requirements.

A simulation application, commonly called a simulator, comprises a mathematical model of the simulated object (a physiological system in this case) and a graphical user interface (GUI). The user interface visually represents the simulated object and its state (as computed by the model) and allows students to control the model via various inputs and controls.

Simulators are used globally to motivate students, enhance their understanding of complex topics, and foster critical thinking and problem-solving skills [1,2]. We have been creating and using simulators designed in various technologies [3] in our classrooms for many years with considerable success [4].

Complex simulators can be confusing for students and thus ineffective without additional explanation. New and effective teaching tools include interactive textbooks that integrate texts with simulators (interactive visualizations driven by models). Students can experiment with the systems and concepts under study using a simulator and thus verify and deepen their understanding. The function of the simulator is explained in the accompanying text and supplemented with suitable scenarios so that students gain maximal utility from the experience. As an example, the interactive textbooks on cardiovascular physiology by Burkhoff and Dickstein are among the first works in this field available on the internet [5] or as an iPad app [6].

Our goal was to develop a technology for the creation of similar teaching materials, but in contrast to Burkhoff [5], our innovations are designed to be platform independent and able to work offline. This technology is free and thus available for anyone to create new interactive teaching materials.

Creating books composed of texts and pictures or animations is technically not difficult. There are several suitable software tools available for this purpose. The challenging task is to include model-driven simulators that allow students to change variables, make predictions, and discover how the system works.

Production of teaching simulators is a demanding and multifaceted task requiring an interdisciplinary team of experts from multiple areas including education, graphics, modeling, and software development.

To begin the process, the educator defines the teaching objectives of the simulator, determines its main design to meet the objectives, and proposes scenarios that the simulator should be able to demonstrate. The educator also defines the demands on the model by determining the (physiological) processes that should be modeled, the parameters that are controlled by the user, and the variables that are displayed in the GUI. The educator also conceptually designs the GUI so that the system and its state is presented in a didactic way.

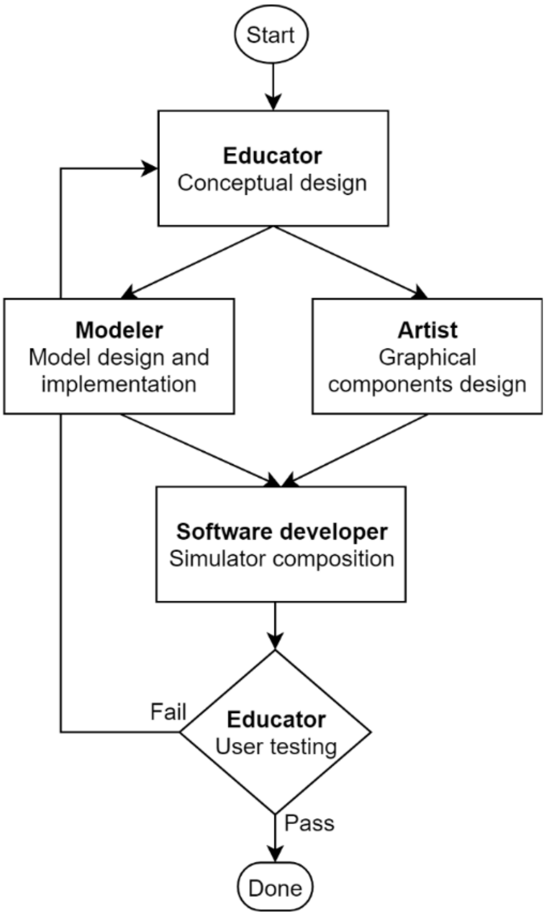
The modeler begins by finding a suitable model in the literature and often combines several models. If a convenient model cannot be found, the modeler derives the model from elementary (eg, physical and physiological) principles. At this point, the modeler implements the model in a programming or modeling language so that its calculation may be run on a computer. If the model is newly developed, it should be verified by comparing its results with credible reference data [7].

The artist designs the visual appearance of the GUI according to the educator's assignment and decides the complete arrangement, including the colors and fonts. The artist also prepares all the drawings including interactive animations that may be linked to respective model variables and thus controlled by the model.

The software developer composes the GUI using the graphical components, stitches it with the model, and produces the final simulator. User controllable elements are bound with the inputs of the model. Output variables from the model are connected with the interactive graphical elements, plots, and other indicators of the GUI. The final application is then deployed on the target platform that can be a Web page or a classical native binary, and eventually, it is embedded into a broader teaching unit.

Finally, the educator tests the simulator in an educational setting and, based on the feedback from students, may decide to iterate a new version of the simulator (Figure 1). The specific roles of the developers, development phases, and several patterns that are useful through the process are described in CoSMos methodology in more detail [8].

**Figure 1.** Process of simulator development.




## Related Work and Approaches

We describe our software engineering efforts toward the development of a lightweight, easy-to-use, platform-independent, open, and standardized framework for creating interactive equation-based simulations for medical education. In this section, we discuss several existing tools related to simulator authoring to explain the reasons why we decided to develop a new toolchain and support our technological decisions.

### *Stand-Alone Executables*

LINDSAY Virtual Human [9] is a 3D interactive model of human anatomy and physiology for medical education. The main focus is on anatomy. This innovative program allows *virtual dissection* and *traveling through the body* visualizations. The physiology models within LINDSAY Virtual Human are mainly implemented using the agent-based approach [10].

*LINDSAY* has been used in several projects [9]. For example, the *anARtonomy* application allows the display and manipulation of anatomical models of several human systems and organs in augmented reality. Similarly, *Zygote 3D Anatomy Atlas & Dissection Lab* is a human anatomy interactive application available for iPad, iPhone, and iPod touch. It is composed of more than 4000 anatomical structures and allows the virtual dissection of any anatomical system. Finally, *Prokaryo* is an interactive simulation application of an *Escherichiacoli* bacterium for the Mac OS. It presents intracellular structures and uses an agent-based physiological model.

The agent-based modeling approach [11] is a useful method applied in many areas where an interaction of multiple autonomous individuals (agents) is simulated. In biology, this method is used, for example, in epidemiological modeling [12], cellular modeling (eg, immune system [13] or tumor growth [14]), or molecular modeling [15]. We focus mainly on physiological models of organs or their systems where traditional modeling based on mathematical relations is predominant. The models we base our simulators on are usually published as systems of mathematical equations, and thus, usage of the equation-based approach is straightforward. Agent-based and equation-based modeling approaches are compared by Parunak et al [16].

*Pulse Physiology Engine* [17] is a comprehensive open-source human physiology model. It is implemented including a solver in C++ and it integrates many physiological systems [18].

Teaching simulators may be based on similar comprehensive physiological models. These models are required for several applications including virtual patient simulators [19], which simulate certain diseases, complex pathological states, or the responses of the body to medications that cannot be described with a simple model.

Subsystems of complex models usually interact in complicated feedback loops, which may be difficult to understand. Therefore, simulators explaining 1 particular system or a body organ are easier to comprehend when based on a smaller model focused only on that system [20], that is, because the single system or the organ is *disconnected* from related subsystems and is not affected by the feedback loops. This method of disconnecting the subsystem is called *Ceteris paribus* (other things equal). Once the student grasps the behavior of an isolated subsystem, more complex and integrated models and simulators may be used to simulate how the subsystems mutually interact.

One important and pioneering tool in simulator production is National Instruments' *LabVIEW*. LabVIEW is designed for control and optimization in engineering with the ability to easily connect to external hardware. Although it was not originally developed for this purpose, it can produce a complete simulator. It allows the developer to implement the model using the block-oriented approach [21] and produce the user interface.

LabVIEW produces an installable executable for multiple operating systems. It has been used for the production of physiological teaching simulators for a long period of time [22] and is currently used for that purpose [23,24], for example, simulators created by AP. Shepherd with LabVIEW are freely available on the Life Science Teaching Resource Community Web page [25]. LabVIEW was also used to control a hardware-based physiological mannequin simulator [26].

In the block-oriented modeling approach used in LabVIEW, the model comprises functional blocks (addition, multiplication, integration, and other more specific blocks). These blocks are connected by their inputs and outputs. The model input values are propagated through the block network and are modified and the output values are calculated [27]. The implementation is visual.

The disadvantage of the block-oriented modeling approach is that the modeler must know which variables are input and which are output before the model implementation. The modeler also has to derive the causality of the model evaluation, that is, the order in which the variables will be successively evaluated [27].

We find the equation-based modeling approach more convenient than the block-oriented approach. The reason for this is that with the equation-based modeling approach, the model components are implemented without any assumption about causality. The causality is resolved by the modeling tool automatically [27] when the model is being translated. Thus, the model components are reusable in different contexts. This makes the approach more convenient for complex models. We strongly prefer tools that offer the advantages of equation-based modeling. These considerations are explained in the *Modelica* section below in more detail.

The GUI is usually composed using predefined components in LabVIEW according to the tutorial [28] and the LabVIEW simulators [25]. These components are designated for the technical domain; thus, the resulting simulator usually has an industrial look.

New interactive animations may be included as a sequence of images using the *picture ring* function as recently described by Jerome [29]. Index of the image in the sequence may be bound to a model variable. This approach is convenient to control the animation with a single variable.

There are some new picture functions available [30] that could facilitate the complex animation production in recent versions of LabVIEW. The LabVIEW NXG Web Module [31,32] allows the programmer to export LabVIEW user interface to a Web browser using an approach technically similar to our own. Both these functionalities make LabVIEW possibly even more useful for teaching simulator production, but we have not found any physiological simulators using these new functions.

In addition to the block-oriented modeling issues discussed above, LabVIEW is a commercial product. This limits its use to individuals who are strongly determined to engage in modeling and simulator development because it requires an investment in a software license. In our experience, employment of commercial products may have the unfavorable effect of discouraging our occasional collaborators from participating in the simulator development.

The iPad version of cardiovascular textbooks by Leisman and Burkhoff [6] is another example of a standalone simulator.

In general, the distribution of a standalone executable may bring installation issues and discourage its use. The executables are platform dependent and must be generated or compiled for each platform separately.

### *Browser-Based Client-Server*

Examples of client-server simulators include JustPhysiology [33] and the Web version of the cardiovascular textbook series by Burkhoff et al [5]. Several technical approaches of how to realize client-server Web-based simulators with special emphasis on Modelica modeling are discussed by Meyer et al [34].

A platform for interactive Modelica content called Modelica University was developed by Tiller and Winkler [35]. Simulators created using this platform are included in the *Modelica by Example* Web-based book by Tiller [36]. The simulator GUI is composed manually in JavaScript.

Žáková and Cech [37] implemented a Web service that runs a Modelica model and implements the JSON-RPC (JavaScript object notation remote procedure call) protocol [38] for the communication with a client application. This application allows individuals to remotely upload, translate, and simulate a model. There are no special tools for the client application production. The client applications using this server back-end may be created using any suitable technology. Several teaching simulators using this technology were created [37], mostly for a course of Control Engineering. We did not find a website of the project nor any realized simulators; therefore, we assume that the project is available for the author's purposes only.

Simulators using the browser-based client-server approach are already platform independent and do not require any installation. Unfortunately, usage in a classroom or during a lecture, where a multitude of students may use the simulator simultaneously, ramps up the computational and bandwidth demands on the server. This can be a major disadvantage.

Latency can be an additional disadvantage of this approach. In fast-paced games, latency greater than approximately 50 ms starts degrading the user experience [39]. Although it is clear that teaching simulators are not as sensitive to latency as real-time gaming, low latency contributes to optimal user experience. For example, when the interaction with the UI produces instant effects, users can continuously move a slider to control a model parameter and quickly observe progressive responses of the system.

The total latency is caused by model evaluation, graphics rendering, and, in the case of client-server architecture, the network communication delay. The internet connection must be fast, and the server must be close enough to achieve low network latency [40].

To solve the issue with an unstable connection in a client-server architecture that would cause the animation to stutter, we would use a caching strategy such as has been done by Brukhoff [5], which would enable us to produce smoothly running animations.

Another disadvantage of this approach is that these simulators require continual connection to the server back-end, and thus, this approach is not suitable for interactive textbooks that need to work offline.

We consider the client-side approach more suitable because renting a cloud service providing enough computing power to serve numerous end users simultaneously and having data centers distributed around the globe to achieve a reasonable round-trip time may become expensive for many individuals and small teams.

On the other hand, the client-server approach is convenient in situations when the client device has insufficient performance to evaluate the model in reasonable time, and the server may deliver the results faster.

### Browser-Based Client Side

To counter the drawbacks of stand-alone and client-server solutions, simulators could be run directly on the client, that is, the browser. Client-side simulators are platform independent. Additional advantages include that they do not require any installation, do not put a heavy load on the server, and can operate on a slower internet connection or fully offline. A major difficulty with this approach is the need to rework any existing platform into a JavaScript codebase. Some pioneering work has already been completed by Wagner [41].

Specifically, we developed a Bodylight(.Net) simulator framework [42], built on a Microsoft Silverlight browser plug-in. In this project, the OpenModelica compiler was extended so that it could translate the model in *C#* language. The OpenModelica runtime and solvers were rewritten manually into the *F#* language. A problem with this approach was that OpenModelica is developing rapidly, and the rewritten runtime must be updated constantly to stay compatible with the rest of the system. The Silverlight plug-in was finally discontinued, and the framework became obsolete and useless.

The *openmodelica-javascript* project by Tom Short [43] extends OpenModelica so that it can compile the models and simulation runtime automatically to JavaScript using *Emscripten* (we also use emscripten). Browser-based simulators comprising text boxes to input parameter values and plots to visualize results can be created easily. As the simulation runtime code is generated automatically, it is much less laborious to keep the system compatible with OpenModelica compared with our obsolete Bodylight(.NET) solution based on Silverlight. The disadvantage is that this project relies only on OpenModelica, and other modeling tools are not supported. Furthermore, more complex GUI elements including sliders and interactive animations are not supported (although it is probably possible to implement them manually). Another concern is that it does not offer any tool to easily compose the GUI, for example, in a drag-and-drop fashion. Owing to the changes in OpenModelica, the openmodelica-javascript project does not currently work with recent versions of OpenModelica (personal communication with Tom Short).

### Related Software Frameworks and Libraries

In addition to the technologies and the software libraries described in the Methods section, there are several other technologies that could possibly be beneficial for use. Here, we discuss some of them.

*BabylonJS* [44] is a free JavaScript 3D engine for games and other 3D visualizations in a Web browser using WebGL. It has several applications for medical e-learning (electronic learning). For example, EducaAnatomia3D is a serious game for human anatomy education [45]. The 3D models may be created in Blender [46] (a free 3D modeling tool) and exported for use in BabylonJS [47]. This combination of BabylonJS and Blender could conveniently extend our toolchain that aims to be based on free tools.

*Tree.js* [48] is another JavaScript engine for 3D visualizations in a Web browser based on WebGL, with applications in medical e-learning [49]. If we decide to include support for 3D graphics, for example, to enable more accurate anatomical visualizations, we could use one of these 2 engines.

Another important game engine and development platform is *Unity* [50], developed by the Unity Technologies company. It allows the innovator to create both 2D and 3D interactive experiences. Unity is available in several versions, and one of the versions is free. Another important aspect of Unity is that it has its own development editor. The main programming language is C#. Unity allows the innovator to build applications for many different platforms. One of the target platforms is WebAssembly [51], which allows the produced simulator to be run in the browser. It is used for the development of serious e-learning games [52-56] in many fields including medical education [57]. We have previously used Unity in combination with Bodylight(.Net) framework in *Surviving Sepsis - a 3D integrative educational simulator* proof-of-concept project (for the Windows Store platform) [58]. Unity is highly convenient to develop complex serious games. If we decide to focus on serious game development, we would consider using Unity, although it is a commercial product.

## Goals for the New Toolchain

There are excellent Web technologies available based on JavaScript, including many useful frameworks and libraries that almost equalize the capabilities of browser and native applications. There are also great modeling tools available. The problem is that the models are deployed in native programming languages (C or C++) so that they cannot be run in a browser.

Our goal was to fill this gap and allow the use of the dedicated modeling tools and enable running the resulting models in the browser without a server-side back-end. On the basis of our previous experience, we have formulated the following requirements on the tool for simulator production:

- *Browser based:* to achieve widespread compatibility and avoid the necessity of an installation process, which could discourage many users.
- *Client side:* so that a multitude of users are able to run the simulator simultaneously without renting expensive computing hardware from a third-party provider.
- *Future proof:* the toolchain should be based on standards, which are unlikely to become obsolete in a reasonable time frame.
- *User friendly:* to enable all those participating in the simulator production to use an appropriate tool for their task so that the work is efficient and pleasing and the resulting simulator is satisfactory.
- *Equation based:* to facilitate model development, their implementation in an equation-based language, preferably Modelica (advantages discussed in the *Methods* section) should be allowed.
- *Open source* and *freely available:* our aim in building the toolchain is to use relevant open source projects and reap the fruits of the hard work of the developers and also contribute back and make the toolkit available for anyone to use under a copyleft license. Open-source licensing should also allow the project to grow in the future and nurture the collaboration between the developers and users. It is especially important for the modeling tool to be freely available. We often collaborate on model development with colleagues from different workplaces, and with a freely available tool, anyone can participate.

There are many tools available, which are useful in the process of the simulator production, and they meet several of the requirements. However, we were unable to find a single tool or toolchain meeting all our requirements. To address this limitation, we developed a toolchain that meets all our specific requirements.

## Methods

We created a new toolchain within the confines of the specified requirements, using the following technologies.

### Modelica

Modelica [59-61] is an equation-based, object-oriented, open standard language for the simulation of complex physical systems.

In the equation-based (acausal) approach, the model is composed of equations that state the relationship between the variables (eg, $x^2 + y^2 = 1$) as opposed to assignments where 1 output variable (on the left-hand side) is assigned a value of an evaluated expression (right-hand side; eg, $y := \mathrm{sqrt}\,(1 - x^2)$) [62]. The causality of the evaluation (order in which equations are evaluated and which variables are calculated from which equation) does not have to be known in the modeling phase. It is derived automatically by the solution tool later [62]. We prefer this approach over block-oriented modeling because, in equation-oriented modeling, the modeler saves the labor of the causality derivation. Moreover, because the model components are implemented without any assumption about causality, they may be easily extended or later reused in various models, where the causality differs [62]. This is very useful in creating reusable component libraries. We base our simulators on models that are usually published as equation systems, and thus, their implementation in the equation-based approach is straightforward.

Below, we illustrate the difference between both approaches with example of harmonic oscillator model described in the equation $m \cdot x'' = -k \cdot x - c \cdot x'$, where $x$ is position, $m$ represents mass, $k$ represents spring constant, and $c$ represents damping constant. Implementation of this model using the block-oriented approach in Modelica (as Modelica supports the block-oriented modeling as well) is illustrated in Figure 2. The code of the same model implemented using the equation-based approach in Modelica is listed in Figure 3. The equation-based (acausal) and block-oriented approaches are compared in several reports [27,62,63].

Owing to its *object-oriented* nature, Modelica scales very well for large complex models [62]. Basic model components (classes) are defined by equations. Components have connectors. More complex components are composed of other components that are connected using their connectors in a visual way. These connections are rendered into additional equations binding the variables of the connected component. Inheritance is also supported and enables significant code reuse. Numerous Modelica libraries of predefined components from many different domains exist. A Modelica library for physiology called Physiolibrary was developed in our laboratory [64,65]. The harmonic oscillator model implemented using Modelica Standard Library is illustrated in Figure 4. The model appearance intuitively explains the modeled system.

Modelica is supported by many different modeling tools, primarily because of the fact that it is an *open standard*. This is important because it allows the use of multiple tools that broaden and expand the modeling capabilities. We use either OpenModelica [66], which is free, or Dassault Systèmes Dymola [67], which is a more advanced, but a proprietary modeling tool.

Modelica is an open-standard language supported by open-source tools and, as such, is ideal for model sharing [68].

**Figure 2.** Harmonic oscillator model implemented using block-oriented approach in Modelica.

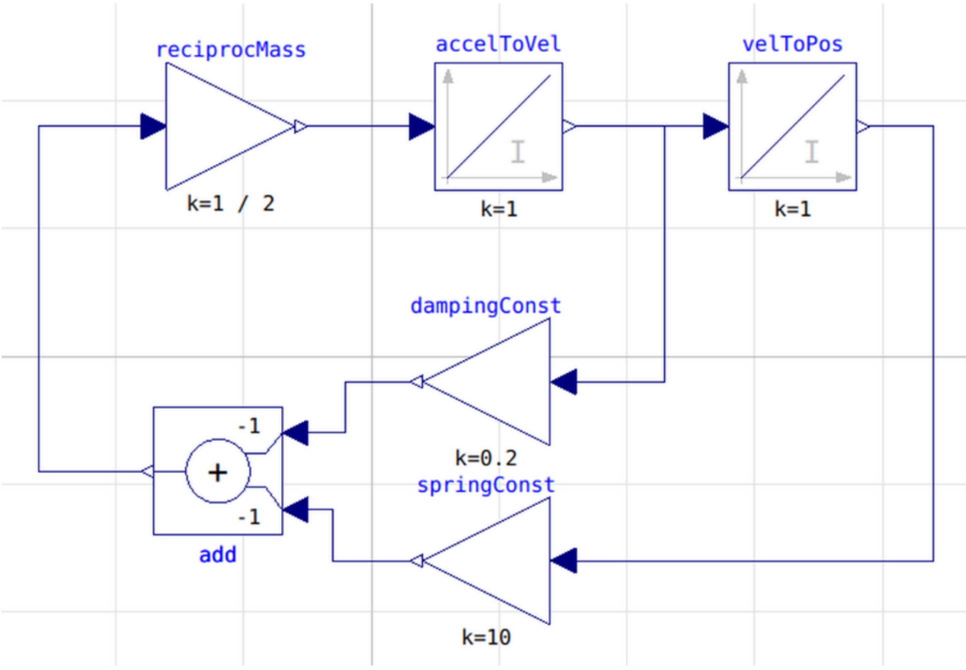
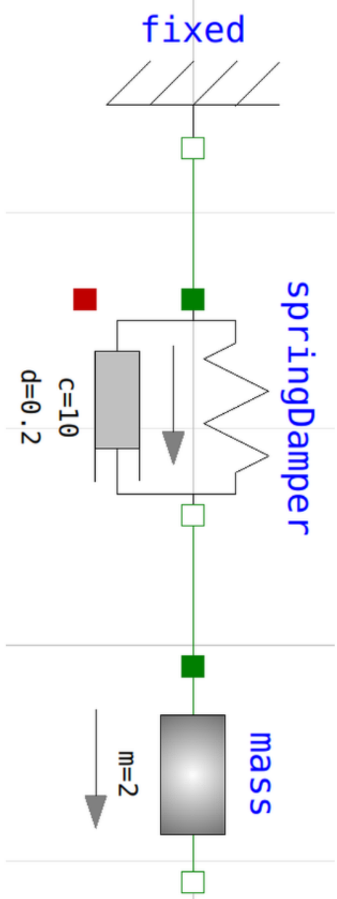



**Figure 3.** Harmonic oscillator model implemented using equation-based approach in Modelica.

```
model OscilatorEq "harmonic oscilator model"
  Real x (start=1, fixed=true) "position";
  Real v (start=0, fixed=true) "velocity";
  parameter Real m = 2 "mass";
  parameter Real c = 10 "spring constant";
  parameter Real d = 0.2 "damping constant";
equation
  m*der(v) = -c*x - d*v;
  der(x) = v;
end OscilatorEq;
```

**Figure 4.** Harmonic oscillator model implemented using Modelica Standard Library.




### Functional Mock-Up Interface

Functional Mock-up Interface (FMI) [69] is a standard for exchanging and cosimulation of dynamic models between different independent tools and applications. Both Dymola and OpenModelica allow for the model to be exported as a Functional Mock-up Unit (FMU). When exported in the mode *FMI for Co-Simulation*, the unit contains a simulation runtime, which takes care of the model calculation and execution. The data exchange with the outside world is restricted to discrete communication points, and between them, the unit is solved independently by the included solver [70].

The FMU can be exported with source code necessary to compile binaries, which implement the Co-Simulation standard. This ability is very advantageous for our purposes as we can compile the source code into a Web language and have full access to the FMI for Co-Simulation features in the browser, which allows us to interact with the model easily.

The FMI standard ensures compatibility with future versions of both OpenModelica and Dymola. Furthermore, the Bodylight.js system can be easily adapted to support other simulation tools implementing the FMI export option.

### JavaScript

JavaScript is a multiplatform, object-oriented, interpreted programming language [71]. It was originally developed by Brendan Eich in 1995. JavaScript is the most notable implementation of the ECMAScript standard. It is supported by all recent important Web browsers [72]. It is widely used today to enable interaction and dynamic behavior on websites.

### JavaScript Libraries

We use the *GrapesJS* open-source Web Builder framework [73] as our HyperText Markup Language (HTML) layout engine. GrapesJS allows us to use the drag-and-drop approach for designing how the simulators appear. There is a set of built-in blocks available to build the app, and it allows for easy customization and production of additional blocks as well. Available configuration panels enable the programmer to edit properties and the behavior of the components on the canvas. The modular design of GrapesJS allows us to hook into its user interface and extend it as a base of the main user experience.

*EaselJS* [74] is a component of the CreateJS toolkit. It allows for easy manipulation of HTML5 Canvas elements and can be used to create games, art, and other graphical experiences. More importantly, for our interests, Adobe Animate natively supports the export of animations to EaselJS. We can either use Adobe Animate to design the animations or EaselJS directly to create original animations.

Finally, *Plotly.js* [75] is a high-level, declarative charting library implemented in JavaScript. It can be used to display many types of charts and graphs.

### WebAssembly

WebAssembly [76] defines a binary instruction format to be executed inside a stack-based virtual machine. Its primary use is to be implemented inside Web browsers, aiming to provide code execution at near-native speeds. The binary format is designed to be efficient with respect to size and load time, reducing the time necessary to transmit and load the code.

WebAssembly is a target of compilation for high-level programming languages such as C or C++, enabling the compilation of existing and new code written in C/C++ to the browser platform. The compiler that facilitates this is the open-source project, Emscripten [77,78], which we use to compile source code inside the FMU. The compiled code can be considered obfuscated to the level similar to those with other binary instruction format representations. The algorithms can be disassembled into a pseudocode and with great effort and investment of time can even be reverse engineered [79]. For most practical purposes, the models can be considered obfuscated when compiled to WebAssembly.

# Results

## Overview

The new Bodylight.js toolchain uses the work of several third-party open-source tools and compilers and other newly written tools. The Web page [80] of this project includes documentation and tutorials. In this section, we describe the complete workflow in more detail and define all the processes involved and the tools used. We focus particular attention on the newly developed *Bodylight.js Composer*, which enables one to create the final simulator by composing the GUI using animations, plots, and control elements in a drag-and-drop style and bind them to the model variables.

## Model Processing

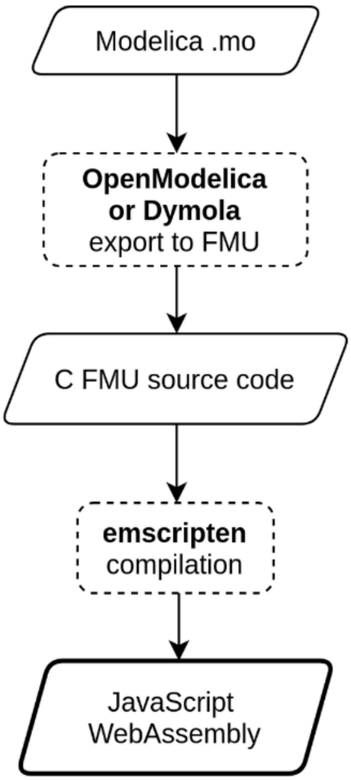
The model workflow is schematically illustrated in Figure 5. The model is written in Modelica, usually inside one of the popular Modelica IDEs such as the open-source OpenModelica or the proprietary and well-established Dymola.

The next step is to export the model into an FMU. Dymola supports the export of source code inside the FMU under their *Source code generation* license [81], and OpenModelica seems to always export the source code inside the FMU. Unfortunately, OpenModelica, as of the date of publication, only allows the export of a Euler solver, which is insufficient for many models.

The FMU then needs to be compiled into WebAssembly and JavaScript. To complete this step, we have prepared a Docker container, which uses Emscripten to automatically compile FMUs into Composer compatible files [82].

The container accepts FMUs generated by Dymola with the *source code generation* license and FMUs generated from OpenModelica on a Linux platform. The requirement for Linux is because of FMUs from OpenModelica being generated with makefiles, which are not only platform dependent but also machine dependent. To facilitate easier environment setup for the OpenModelica part of this workflow, we have prepared another Docker container, which uses OpenModelica to export FMUs automatically [83].

**Figure 5.** Model workflow—Modelica model is exported from OpenModelica or Dymola to a Functional Mock-up Unit (FMU), which is compiled using emscripten into WebAssembly and JavaScript.



### Animation Processing

We use Adobe Animate [84] in the process of designing the interactive animations. Adobe Animate supports native export to HTML5 and JavaScript using the library EaselJS. This process is fairly painless as all the post processing of the exported JavaScript code is handled by the composer. Furthermore, the user can opt to write all EaselJS directly without the need to use the Adobe product.

### Bodylight.js Composer

*Bodylight.js Composer* is the main development focus of this project. Composer is a single-page application that can easily bring together models, animations, and control elements.

The core of Composer is built on React, an established and very popular JavaScript library. The HTML layout engine is provided by GrapesJS, around which the rest of the application was shaped. Composer allows the user to easily design an interactive HTML simulator. There are also several input and output widgets available. The *range* widget handles the control of the model variables. The *chart* widget uses plotly.js to display an output from the model in interactive charts. Toggle widgets and buttons can control Boolean values, and labels are used to display values.

The Composer is equipped with *actions* that are user-generated snippets of JavaScript code and can be attached to events of other widgets. For example, one of the prefilled actions is *reset model* with a model as a parameter. Users can attach *actions* to an *onclick* event of the *button* widget and select the appropriate model to reset. The *animate* widget is used to import complex animations created in Adobe Animate. These can contain continuously playing animations, whose speed and direction can be controlled by any model variable. Furthermore, it can control positional animations, where the timeline position is directly controlled by a model variable.

Users can also save and open shareable project files. The final export from the Composer is a stand-alone HTML file, containing the JavaScript and WebAssembly code. Composer workflow is depicted in Figure 6 and the composer itself in Figure 7.

We recommend readers to view the video tutorials on the Bodylight.js Web page [80] to get a better understanding of how Bodylight.js Composer works. The simple *Bouncing Ball* video on the main page demonstrates how to build a simulator comprised a simple interactive animation, a plot, a slider, and a reset button. It is also included in the Multimedia Appendix 1. There are 2 additional video tutorials available in the *Documentation* section of the Web page. The *Simple Project* tutorial [85] demonstrates how to build a simulator composed of sliders, plots, and a reset button. It is logically sectioned into different episodes. The steps are also explained below the tutorial videos. The second tutorial available is the more elaborate *Physiological Application* tutorial [86]. This tutorial demonstrates several advanced features for building a real simulation application of pressure-volume cardiac loops depicted in Figure 8. The additional features explained include creating a parametric plot, adding a start-stop button, controlling model parameters by events, and applying user functions on model variables. For example, descriptions on how to round the values or display string messages and other features are discussed. The translated models and the Composer project files are included in both tutorials.

**Figure 6.** Composer workflow scheme—HTML (Hypertext Markup Language) layout is created, animations are loaded, the model is loaded, model and animation variables are bound, and control elements and plots are added and bound with model variables.

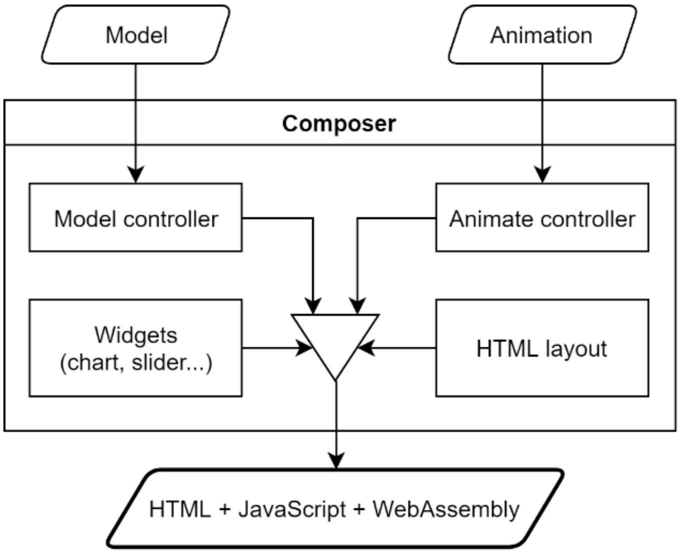


**Figure 7.** The Glomerulus application page inside Bodylight.js Composer.

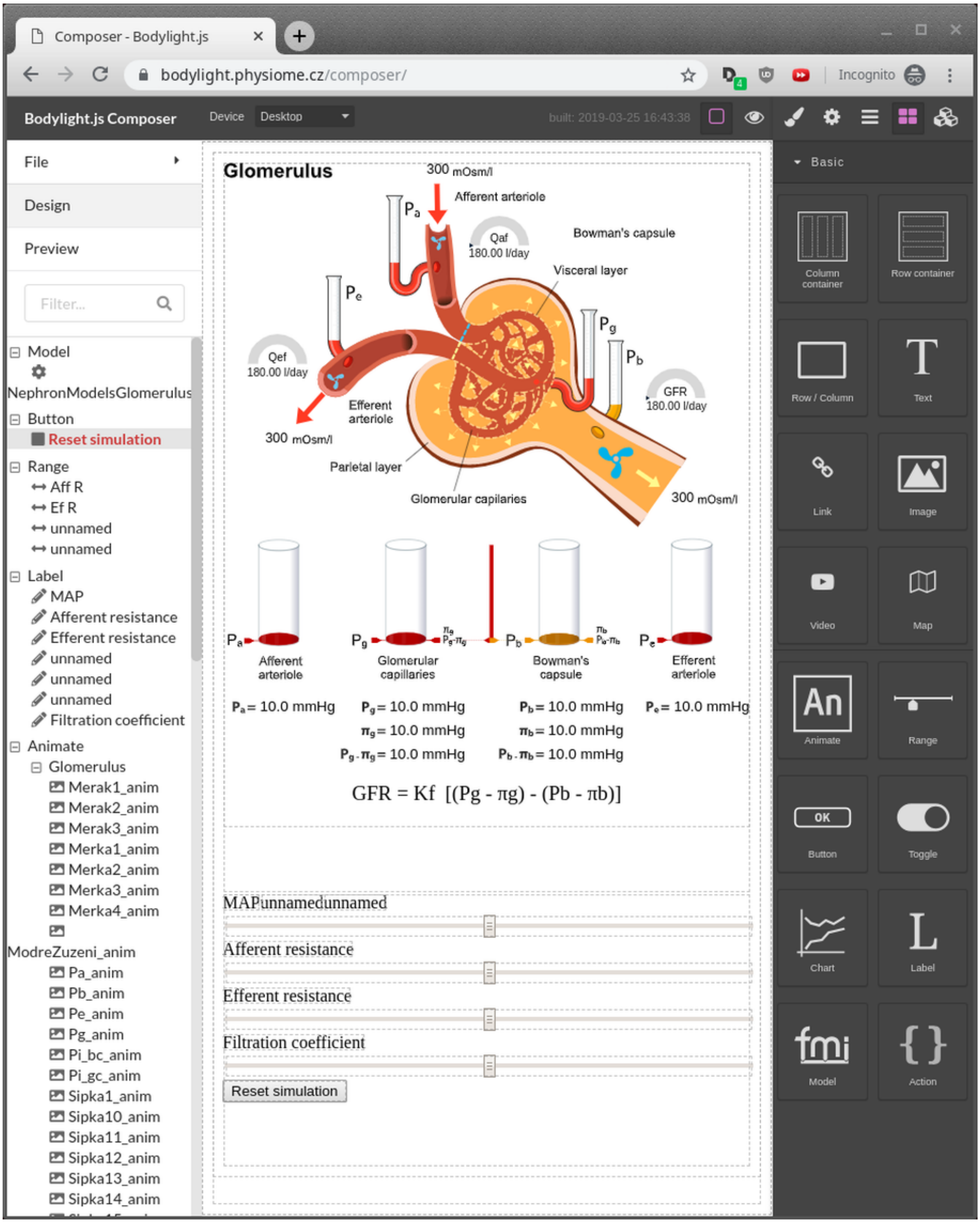
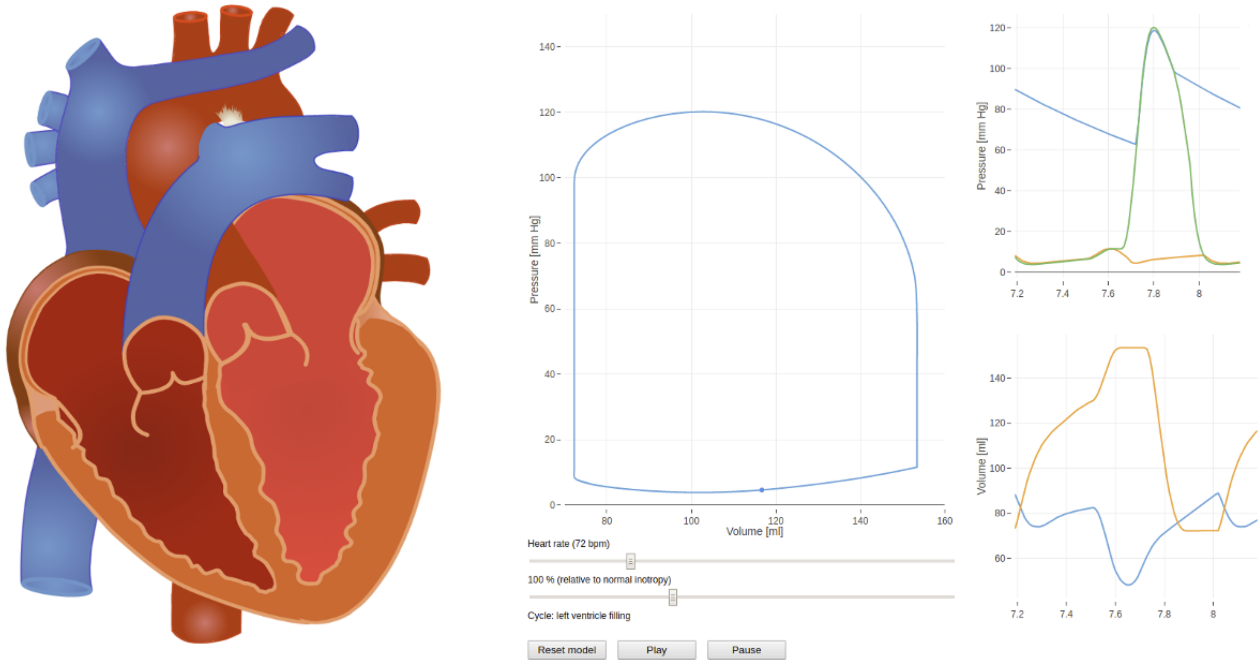


**Figure 8.** Pressure-volume loop simulator is a result of the more advanced tutorial.




## Original Simulators

Several simulators were created using Bodylight.js. In this section, one of the simulators is presented to demonstrate the capabilities of Bodylight.js. First, some basic physiology is introduced, and then the simulator is described, and the features of Bodylight.js are highlighted.

### *Nephron Simulator*

The main purpose of the kidney is to produce urine and control its composition. The functional unit of the kidney is the nephron. There are approximately 2 million nephrons in a pair of kidneys. The nephron is composed of the glomerulus and a system of tubules. The glomerulus is a network of capillaries. The blood is filtered across the capillary walls; thus, primary urine (filtrate) is produced (approximately 180 L/day). The filtrate then flows through the system of consequent tubules, each having a slightly different function, where the water and specific solutes are reabsorbed so that the appropriate amount of urine with the required composition to maintain homeostasis is produced and excreted.

These processes are explained visually by the simulator. For simplicity, the stimulator only focuses on water and sodium. The simulator is available online [87] and is attached as the Multimedia Appendix 2. Multimedia Appendix 2 also includes the additional required libraries for offline use. A more detailed description of the simulator including its didactic objectives, models, and implementation is beyond the scope of this paper. More information is available in our recent work [88].

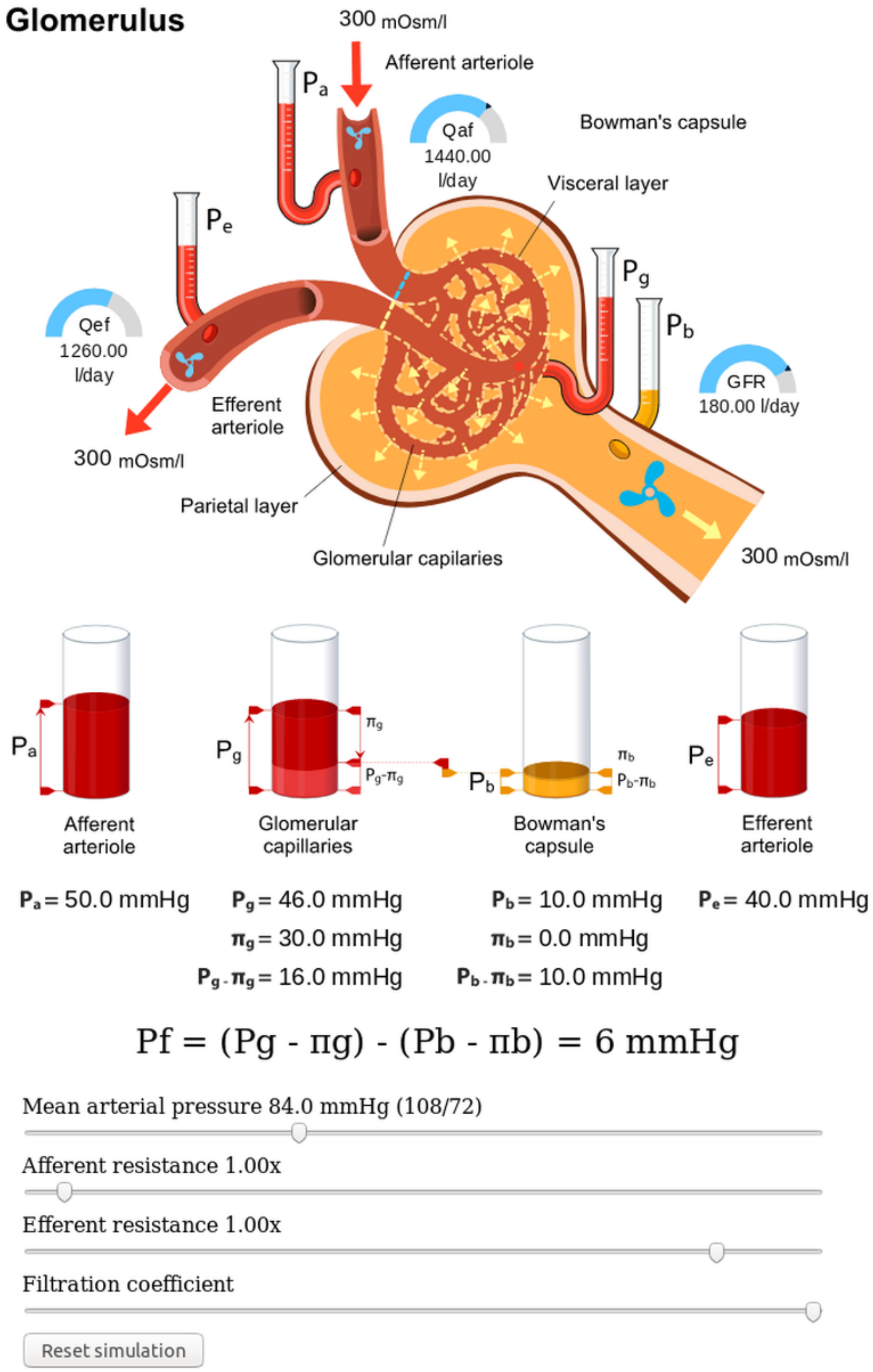
#### Glomerulus Screen

Figure 9 depicts the glomerulus screen. Resistances of the arterioles, mean arterial pressure, and the filtration coefficient are controlled with the sliders. Changes of these parameters affect pressures and flows in the system. Pressures are depicted by the liquid-column gauge and flows through the tubules by the speed of the propellers and the half-circle indicators (normal values are marked by a tick). Flow through the vessel walls is shown with the width of the dashed moving arrows. The hydrostatic and oncotic pressures are also indicated by the cylinders below the glomerulus. All the indicators are interactive and react to changes in values of the model variables.

Note that the numeric values of the flow indicators and the pressure values in the formulae are also controlled by the model. There is also a *Reset simulation* button which sets all the parameters to default values.

There is no time evolution in the model, and everything is computed in the initialization phase. The simulator operates in the so-called *One shot* regime. This means every time parameters are changed by the sliders, the model is automatically recalculated and the GUI is updated accordingly. The same also holds for the next screen.

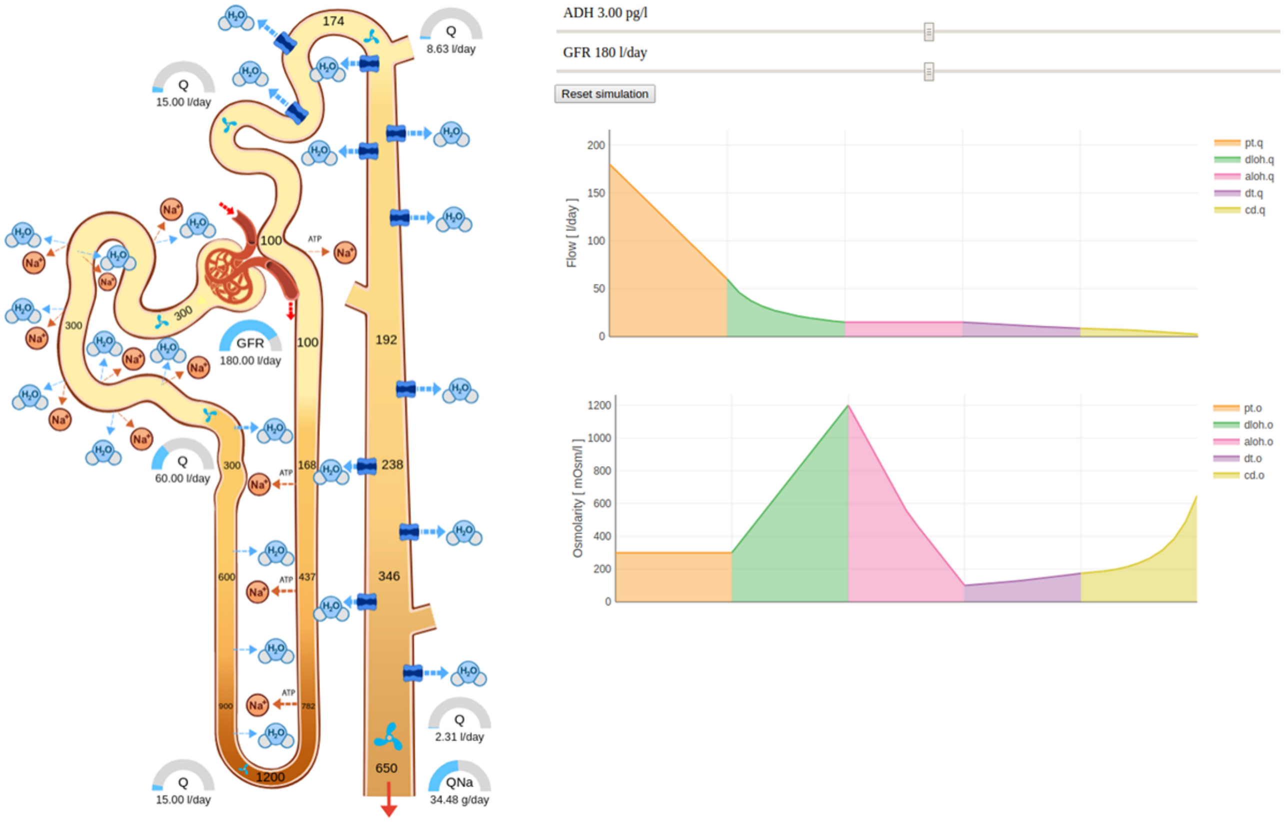
**Figure 9.** Glomerulus. Pressure is visualized by the liquid-column indicators and flows by the propellers and half-circle indicators. The red arrows symbolize the blood flow direction, and yellow arrows represent urine flow direction. GFR: glomerular filtration rate.




## Complete Nephron Screen

The simulator of the complete nephron is shown in Figure 10. The glomerular filtration rate and the antidiuretic hormone (ADH) parameters are controlled with the sliders. Filtrate flow rate and the sodium mass flow rate are visualized by the half-circle indicators. Osmolarity (proportional to solute molar concentration) of the filtrate is determined by the numbers and the lightness gradient inside the tubules. Water and sodium flow through the tubule walls are determined by the width of the blue- and orange-dashed moving arrows. The amount of ADH and the consequent tubule water permeability is visualized by the varying width of the blue water channels in the last section of the tubule.

**Figure 10.** Complete nephron. Half-circle gauges show urine flow rate, the lightness gradient inside the tubules visualizes osmolarity, dashed blue and orange moving arrows visualize water and sodium flow through the tubule walls and the blue channels indicate tubule water permeability. Flow and osmolarity are plotted in the charts (individual sections in a different color). ADH: antidiuretic hormone; GFR: glomerular filtration rate.

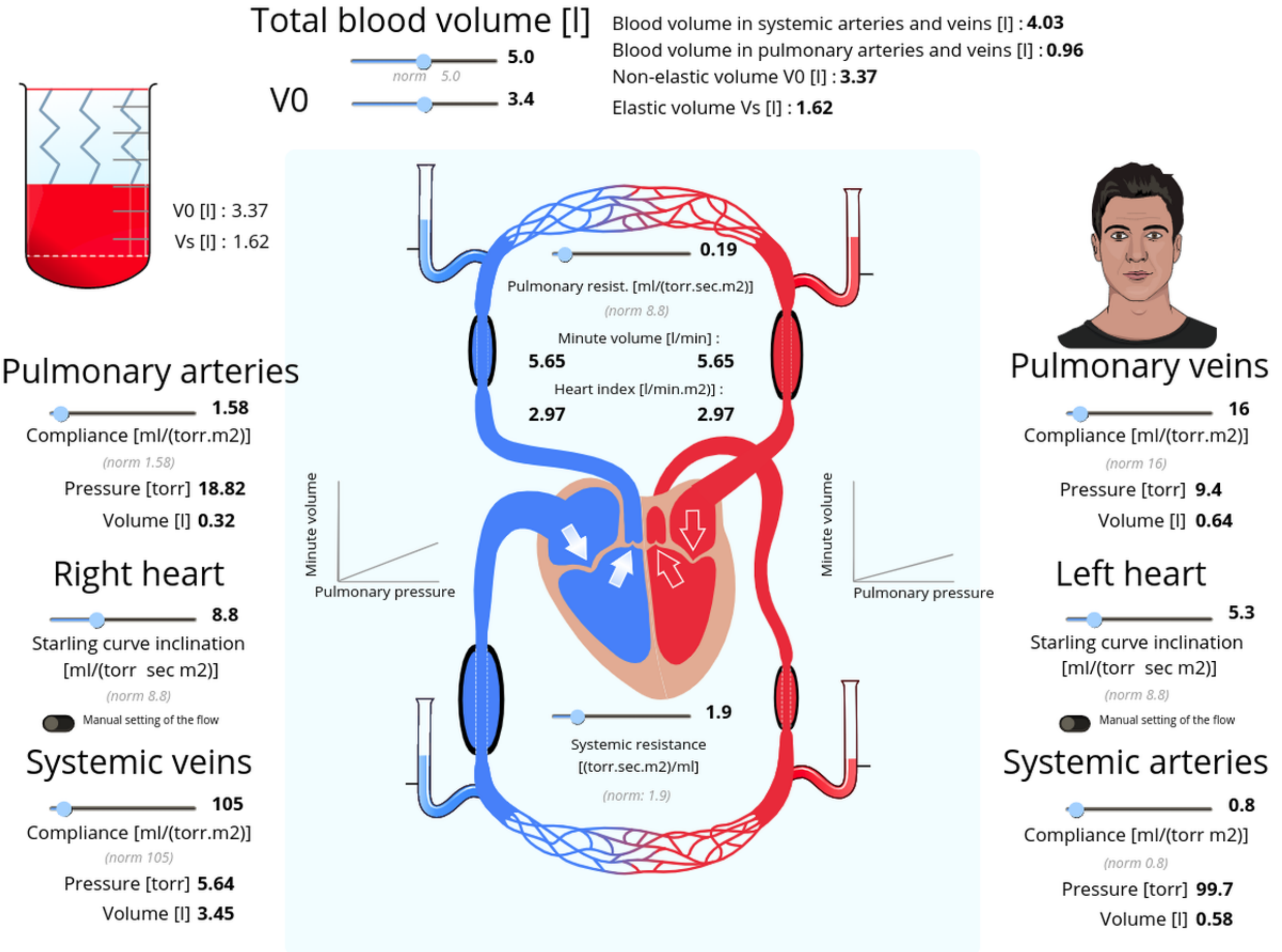


Flow and osmolarity are functions of distance along the nephron tubules and are plotted on the chart. The individual sections, whose variables are discretized using an array in the model, are highlighted in different colors for easier identification. The legend, the labels, the line color, and the width as well as several other plot properties are also adjusted.

### Blood Circulation

The blood circulation simulator (Figure 11), available online [89], is another example illustrating the use of Bodylight.js Composer. The simulator depicts 2 general blood vessel circuits: (1) a large systemic circuit distributing blood to the body and (2) a small pulmonary circuit running through the lungs. The circuits are connected through the heart. The simulator is based on a basic physical model with the default parameter values fitted to clinical values to provide meaningful results. Students can modify various model parameters: (1) the compliance of systemic and pulmonary arteries and veins (reciprocal to elasticity), (2) the slope of the Starling curve reflecting cardiac contractility and cardiac output, (3) the total and nonelastic (unstressed, V0) blood volume, and (4) the pulmonary and systemic resistances. The simulator is supplemented with animated figures; therefore, the system response on external perturbation can be followed graphically (intuitively) as well as through the numeric values calculated by the model running in the background. Note that the human face animation is also controlled by arterial pressure, with facial changes in expression and color representing low, normal, or high pressures.

**Figure 11.** Blood circulation simulator. Pressures are visualized by liquid-column gauges, blood volumes are depicted by the width of the sections of blood vessels with black borders, blood flow rates are visualized by the frequency of blinking arrows. Changes in arterial pressure are also reflected by animated changes in facial expression and color.




### *Pressure-Volume Loop Simulator*

A cardiovascular simulator (Figure 8), available on the Web page [90], displays the pressure-volume loops of the left ventricle. It is based on our Modelica implementation [91] of the model by Burkhoff and Tyberg [92]. The model-controlled image displays atrial and ventricular filling, as well as valve opening and closing during the cardiac cycle. Students can pause the simulation and track the names of the cardiac cycle phases, atrial and ventricular pressures and volumes, pulmonary artery and aortic pressures, and the current point on the pressure-volume loop. The multimedia tutorial on the Web page [80] describes how to create this simulator from the original Modelica model. The simulator was mainly developed for the purpose of the tutorial. If it was intended for education, more plots and sliders would be required.

## Discussion

### Principal Results

We created the Bodylight.js toolchain to facilitate the development of interactive simulators based on Modelica models. In this report, we focus on describing an important new component of the toolchain (Bodylight.js Composer), which enables the creation of browser-based simulators. More information about the toolchain and its use is available on the Web page [80].

The goals for the toolchain were addressed. Importantly, both Bodylight.js Composer and the simulators it produces are *browser based* and *client side*. The system runs in the browser without the need of any server-side back-end. It is also possible to distribute a standalone platform-independent HTML file and run it in the browser without an internet connection. This is enabled by the use of Web technologies such as JavaScript, WebAssembly, and HTML. We believe it will be *future-proof* (unlikely to become obsolete), as it is mainly based on open-standard technologies accepted and implemented by every major software vendor. If any of the applications in the toolchain are discontinued, it should be easy to replace them with another tool. We also find it to be *user friendly*. The models are implemented in the *equation-based Modelica* language using a modeling environment of the modeler's choice, for example, OpenModelica or Dymola. Animations are created in a professional industry standard tool, Adobe Animate. The Composer uses the drag-and-drop technique to visually compose the simulator, and it is distributed under the General Public License 3.0 *open source* license and is *freely available*. Therefore, it is available for anyone to use and implement within their open-source projects. To our knowledge, no other tool exists, which meets all our requirements.

This approach brings together the domains of modeling, Web technologies, and graphical design, which supports better interdisciplinary cooperation of teachers, modelers, software developers, and graphic designers.

The Bodylight.js Composer is a self-contained client-side application, and anything submitted does not leave the user's device; thus, there are no particular security or privacy issues.

Several simulators were created in Bodylight.js, and it proved to be a convenient solution fulfilling the needs of the designers. It would be extremely time consuming to implement these applications without this versatile toolchain. The new Nephron, Blood Circulation, and Pressure-Volume Loop simulators were presented in this paper to demonstrate the capabilities of Bodylight.js. The Nephron simulator was recently used in didactic classes of pathological physiology for medical students and was very well received by both the students and the teacher.

### Comparison With Previous Work

The client-side approach has several advantages over the client-server solutions available today. First, the simulator is not bound by the computational limitations of the server. The scaling problem is bypassed by avoiding the necessity of any server-side computations. Instead, we simply serve a Web page, which can be hosted on any Web server. Second, the client-side solution does not require a low-latency connection to the server. Thus, the client-side approach avoids the need for expensive Web hosting services.

E-learning and distance learning are becoming increasingly important in the world and particularly important in developing countries, where teachers are not easily accessible by many potential students [93]. Furthermore, because reliable internet connections may not be available for many people [94] and a round-trip latency is often high [95] in developing countries, the client-side solution could be more suitable here.

### Limitations

The OpenModelica export to FMU for Co-Simulation is currently limited to the Euler solver, without the option to include any of the other solvers available in the OpenModelica compiler. Euler is the simplest solver available, and its numerical performance is generally poor compared with more advanced solvers included in OpenModelica. Thus, the export from OpenModelica is currently not viable for many models. We are often forced to export the FMU from the proprietary Dymola, and this will continue at least until the issue with OpenModelica FMU export has been resolved.

The Adobe Animate, which is used for creation of the animations, is a commercial tool. JavaScript code describing the animations using the EaselJS library may be written by hand as an alternative.

In situations of a computationally complicated model or multiple plots or animations in the simulator, the performance drop becomes noticeable. The slower frame rate or model update rate does not make the simulator useless, but the user experience is reduced.

Bodylight.js relies on relatively new Web technologies; therefore, only browsers after late 2017 are able to run our simulators. However, because running older browsers is a security risk, most browser vendors have already switched to an automatic update system.

### Future

We will extend OpenModelica so that it can use advanced solvers in the FMU for Co-Simulation. We plan to optimize the model calculation and graphics rendering to achieve higher frame rates. The composer is modular in design; therefore, we are planning to add support for new libraries, such as different charting libraries or even other animation libraries, and external code contribution is welcomed.

### Conclusions

The new toolchain facilitates the production of teaching simulators not only in physiology but also in other fields where the behavior of systems can be described by mathematical equations, including biology, physics, chemistry, engineering, and economics. Furthermore, the technology is not limited to education or academia. Anyone with the ability to model their system, whatever it may be, can use Bodylight.js to visually explain it to any interested party.

New generations of electronic textbooks combining texts with images, animations, multimedia, and interactive model-driven simulators are emerging. These textbooks allow for experimentation with the simulation of the particular systems being taught, which contributes to a deeper understanding of the topics of interest.

We recommend that the teaching materials be developed as platform independent in-browser applications, which do not require installation and can operate without an internet connection. Bodylight.js fulfills all these requirements and is free and available for anyone to use, which can only help to increase its impact. We hope that this project will help people better understand a multitude of diverse systems.

### Acknowledgments

This work was supported by the TRIO MPO FV20628, FV30195, PROGRESS Q26 grants, and the Creative Connections company

### Conflicts of Interest

Creative Connections aims to start business on consulting, providing animations, models, or complete simulators based on the Bodylight.js platform.

Multimedia Appendix 1

Bodylight.js Composer Bouncing Ball tutorial video.
[MP4 File (MP4 Video), 123237 KB - jmir_v21i7e14160_app1.mp4 ]

Multimedia Appendix 2
Nephron simulator.
[ZIP File (Zip Archive), 13740 KB - jmir_v21i7e14160_app2.zip ]

## References

1. Fischer Q, Sbissa Y, Nhan P, Adjedj J, Picard F, Mignon A, et al. Use of simulator-based teaching to improve medical students' knowledge and competencies: randomized controlled trial. J Med Internet Res 2018 Dec 24;20(9):e261 [FREE Full text] [doi: 10.2196/jmir.9634] [Medline: 30249587]
2. Pennaforte T, Moussa A, Loye N, Charlin B, Audétat MC. Exploring a new simulation approach to improve clinical reasoning teaching and assessment: randomized trial protocol. JMIR Res Protoc 2016 Feb 17;5(1):e26 [FREE Full text] [doi: 10.2196/resprot.4938] [Medline: 26888076]
3. Kofránek J, Mateják M, Privitzer P. KOFRLAB: Laboratory of Biocybernetics and Computer Aided Learning. 2010. MEFANET Report 03: Web Simulator Creation Technology URL: http://www.physiome.cz/references/mefanetreport3.pdf [accessed 2019-04-05]
4. Kofranek J, Matousek S, Rusz J, Stodulka P, Privitzer P, Matejak M, et al. The atlas of physiology and pathophysiology: web-based multimedia enabled interactive simulations. Comput Methods Programs Biomed 2011 Nov;104(2):143-153. [doi: 10.1016/j.cmpb.2010.12.007] [Medline: 21232813]
5. Burkhoff D, Dickstein ML. Harvi. URL: http://harvi.online/ [accessed 2019-03-20] [WebCite Cache ID 7717nqxOO]
6. Leisman S, Burkhoff D. Use of an iPad app to simulate pressure-volume loops and cardiovascular physiology. Adv Physiol Educ 2017 Sep 1;41(3):415-424 [FREE Full text] [doi: 10.1152/advan.00204.2016] [Medline: 28679580]
7. Kurtz TW, DiCarlo SE, Pravenec M, Ježek F, Šilar J, Kofránek J, et al. Testing computer models predicting human responses to a high-salt diet. Hypertension 2018 Dec;72(6):1407-1416. [doi: 10.1161/HYPERTENSIONAHA.118.11552] [Medline: 30571226]
8. Andrews PS, Polack FA, Sampson AT, Stepney S, Timmis J. University of York. 2010. The CoSMoS Process, Version 0.1: A Process for the Modelling and Simulation of Complex Systems URL: http://www-users.cs.york.ac.uk/~susan/bib/ss/nonstd/tr453.pdf [accessed 2019-04-06]
9. Jacob C, Hallgrimsson B, Coderre S, Jamniczky H. LINDSAY Virtual Human. URL: http://www.lindsayvirtualhuman.com [accessed 2019-04-27] [WebCite Cache ID 77wndusCm]
10. Jacob C, von Mammen S, Davison T, Sarraf-Shirazi A, Sarpe V, Esmaeili A, et al. LINDSAY virtual human: multi-scale, agent-based, and interactive. Adv Intell Model Simul 2012;422:327-349 [FREE Full text] [doi: 10.1007/978-3-642-30154-4-14]
11. Wilensky U, Rand W. An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems With NetLogo. Cambridge, MA, USA: The MIT Press; 2015.
12. Marshall BD, Galea S. Formalizing the role of agent-based modeling in causal inference and epidemiology. Am J Epidemiol 2015 Jan 15;181(2):92-99 [FREE Full text] [doi: 10.1093/aje/kwu274] [Medline: 25480821]
13. Fachada N, Lopes V, Rosa A. Agent-Based Modelling and Simulation of the Immune System: A Review. In: Proceedings of the 13th Portugese Conference on Artificial Intelligence. 2007 Presented at: EPIA'07; December 3-7, 2007; Guimaraes, Portugal.
14. Zhang L, Athale CA, Deisboeck TS. Development of a three-dimensional multiscale agent-based tumor model: simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. J Theor Biol 2007 Jan 7;244(1):96-107. [doi: 10.1016/j.jtbi.2006.06.034] [Medline: 16949103]
15. Azimi M, Bulat E, Weis K, Mofrad MR. An agent-based model for mRNA export through the nuclear pore complex. Mol Biol Cell 2014 Nov 5;25(22):3643-3653 [FREE Full text] [doi: 10.1091/mbc.E14-06-1065] [Medline: 25253717]
16. van Dyke PH, Savit R, Riolo RL. Agent-Based Modeling vs Equation-Based Modeling: A Case Study and Users' Guide. In: Proceedings of the First International Workshop on Multi-Agent Systems and Agent-Based Simulation. 1998 Presented at: MABS'98; July 4-6, 1998; Paris, France p. 10-25. [doi: 10.1007/10692956_2]
17. Pulse: Physiology Engine. URL: http://physiology.kitware.com [accessed 2019-04-27] [WebCite Cache ID 77wyMyEpf]
18. Bray A, Webb JB, Enquobahrie A, Vicory J, Heneghan J, Hubal R, et al. Pulse physiology engine: an open-source software platform for computational modeling of human medical simulation. SN Compr Clin Med 2019 Mar 27;1(5):362-377 [FREE Full text] [doi: 10.1007/s42399-019-00053-w]
19. Kofránek J, Kulhánek T, Mateják M, Ježek F, Šilar J. Integrative Physiology in Modelica. In: Proceedings of the 12th International Modelica Conference. 2017 Presented at: Modelica'17; May 15-17, 2017; Prague, Czech Republic p. 15-17. [doi: 10.3384/ecp17132589]
20. Kofránek J, Vu LD, Snáselová H, Kerekes R, Velan T. GOLEM--multimedia simulator for medical education. Stud Health Technol Inform 2001;84(Pt 2):1042-1046. [doi: 10.3233/978-1-60750-928-8-1042] [Medline: 11604890]

21. Lipovszki G, Aradi P. Simulating complex systems and processes in LabVIEW. J Math Sci 2006 Feb;132(5):629-636 [FREE Full text] [doi: 10.1007/s10958-006-0007-z]
22. Kiel JW, Shepherd AP. A graphic computer language for physiology simulations. Comput Life Sci Educ 1988;5(7):49-56 [FREE Full text]
23. Lin SL, Guo NR, Chiu CC. Modeling and simulation of respiratory control with LabVIEW. J Med Biol Eng 2012;32(1):51-60 [FREE Full text] [doi: 10.5405/jmbe.829]
24. Cole RT, Lucas CL, Cascio WE, Johnson TA. A LabVIEW model incorporating an open-loop arterial impedance and a closed-loop circulatory system. Ann Biomed Eng 2005 Nov;33(11):1555-1573. [doi: 10.1007/s10439-005-7785-1] [Medline: 16341923]
25. Life Science Teaching Resources Community. URL: https://www.lifescitrc.org [accessed 2019-03-20]
26. Samosky JT, Nelson DA, Wang B, Bregman R, Hosmer A, Mikulis B, et al. BodyExplorerAR: Enhancing a Mannequin Medical Simulator With Sensing and Projective Augmented Reality for Exploring Dynamic Anatomy and Physiology. In: Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction. 2012 Presented at: TEI'12; February 19-22, 2012; Kingston, Ontario, Canada p. 263-270. [doi: 10.1145/2148131.2148187]
27. Kofránek J, Mateják M, Privitzer P, Tribula M. Laboratory of Biocybernetics and Computer Aided Learning. 2008. Causal or Acausal Modelling: Labour for Humans or Labour for Machines URL: http://www.physiome.cz/references/tcp2008.pdf
28. National Instruments. 2018. Tutorial: User Interface URL: http://www.ni.com/tutorial/7568/en/ [accessed 2019-04-22] [WebCite Cache ID 77p13K4Nn]
29. Jerome J. Virtual Instrumentation Using LabVIEW. New Delhi, India: Phil Learning Pvt Ltd; 2010.
30. National Instruments. 2018. Using the 2D Picture Control URL: http://zone.ni.com/reference/en-XX/help/371361R-01/lvconcepts/using_the_picture_indicator [accessed 2019-05-04] [WebCite Cache ID 787dph3MN]
31. National Instruments. What Is the LabVIEW NXG Web Module? URL: http://www.ni.com/cs-cz/shop/electronic-test-instrumentation/add-ons-for-electronic-test-and-instrumentation/what-is-labview-nxg-web-module.html [accessed 2019-05-04] [WebCite Cache ID 787eTRcZH]
32. WebVIs: Developing Your Web-Based User Interface. URL: http://www.webvi.io [accessed 2019-05-11]
33. Just Physiology. URL: http://justphysiology.com [accessed 2019-03-20]
34. Meyer zu Eissen S, Stein B. Realization of web-based simulation services. Comput Ind 2006 Apr;57(3):261-271 [FREE Full text] [doi: 10.1016/j.compind.2005.12.007]
35. Tiller MM, Winkler D. modelica.university: A Platform for Interactive Modelica Content. In: Proceedings of the 12th International Modelica Conference. 2017 Presented at: Modelica'17; May 15-17, 2017; Prague, Czech Republic. [doi: 10.3384/ecp17132725]
36. Tiller MM. Modelica University. 2019. Modelica by Example URL: http://mbe.modelica.university [accessed 2019-04-05]
37. Žáková K, Cech M. Design of Control Education Interactive Examples via Web Service for OpenModelica. In: Proceedings of the 13th APCA International Conference on Automatic Control and Soft Computing. 2018 Presented at: IEEE'18; June 4-6, 2018; Ponta Delgada, Portugal p. 242-246. [doi: 10.1109/CONTROLO.2018.8514288]
38. JSON-RPC Working Group. JSON-RPC. 2004. JSON-RPC 2.0 Specification URL: http://www.jsonrpc.org/specification [accessed 2019-04-29] [WebCite Cache ID 77zeitMDR]
39. Raaen K, Grønli TM. Latency Thresholds for Usability in Games: A Survey. In: Proceedings of the Norwegian Informatics Conference. 2014 Presented at: NIK'14; November 17-19, 2014; Bergen, Norway.
40. McManus JP, Day TG, Mailloux ZJ. Digital WPI: Worcester Polytechnic Institute. 2019. The Effects of Latency, Bandwidth, and Packet Loss on Cloud-Based Gaming Services URL: https://digitalcommons.wpi.edu/cgi/viewcontent.cgi?article=6326&context=iqp-all [accessed 2019-05-06]
41. Wagner G. Sim4edu.com: Web-Based Simulation for Education. In: Proceedings of the Winter Simulation Conference. 2017 Presented at: IEEE'17; December 3-6, 2017; Las Vegas, Nevada, USA p. 4240-4251. [doi: 10.1109/wsc.2017.8248130]
42. Silar J, Kofranek J. Physiological model creation and sharing. Eur J Biomed Inform 2011;7(1):55-58 [FREE Full text] [doi: 10.24105/ejbi.2011.07.1.10]
43. Short T. GitHub Inc. 2019. OpenModelica-Javascript URL: http://github.com/tshort/openmodelica-javascript [accessed 2019-04-28] [WebCite Cache ID 77yEwGvel]
44. BabylonJS. URL: http://www.babylonjs.com [accessed 2019-05-05] [WebCite Cache ID 7891VBQgM]
45. Batista AV, Lemos RR, Rudolph CM, Bueno BS, Fiuza PJ, Conceicao KR, et al. Design of A Web3D Serious Game for Human Anatomy Education. In: Handbook of Research on Immersive Digital Games in Educational Environments. Pennsylvania, USA: IGI Global; 2017:286-611.
46. Blender. URL: http://www.blender.org [accessed 2019-05-05] [WebCite Cache ID 7894dRF6M]
47. BabylonJS Documentation. URL: http://doc.babylonjs.com/resources/blender [accessed 2019-05-05] [WebCite Cache ID 7893Xwifk]
48. ThreeJS. URL: http://threejs.org [accessed 2019-05-05] [WebCite Cache ID 789AqFViY]
49. Kasinathan V, Mustapha A, Nur FA, Zainal AA. Three-dimensional e-learning application for anatomy and physiology of brain. Int J Integr Eng 2018 Nov 1;10(6):144-148 [FREE Full text] [doi: 10.30880/ijie.2018.10.06.020]
50. Unity. URL: https://unity.com/ [accessed 2019-05-04]

51. Trivellato M. Unity. 2018. Unity Blog: WebAssembly is Here! URL: https://blogs.unity3d.com/2018/08/15/webassembly-is-here/ [accessed 2019-05-04] [WebCite Cache ID 787knmtIA]
52. Horachek D. Creating eLearning Games With Unity. Birmingham, England: Packt Publishing; 2014.
53. Zarzuela MM, Pernas FJ, Martínez LB, Ortega DG, Rodríguez MA. Mobile serious game using augmented reality for supporting children's learning about animals. Procedia Comput Sci 2013;25:375-381 [FREE Full text] [doi: 10.1016/j.procs.2013.11.046]
54. Coelho A, Kato E, Xavier J, Gonçalves R. Serious Game for Introductory Programming. In: Proceedings of the Second International Conference on Serious Games Development and Applications. 2011 Presented at: SGDA'11; September 19-20, 2011; Lisbon, Portugal p. 61-71. [doi: 10.1007/978-3-642-23834-5_6]
55. George AK, McLain ML, Bijlani K, Jayakrishnan R, Bhavani RR. A Novel Approach for Training Crane Operators: Serious Game on Crane Simulator. In: Proceedings of the Eighth International Conference on Technology for Education. 2016 Presented at: IEEE'16; December 2-4, 2016; Mumbai, India p. 116-119. [doi: 10.1109/T4E.2016.030]
56. Boada I, Rodriguez-Benitez A, Garcia-Gonzalez JM, Olivet J, Carreras V, Sbert M. Using a serious game to complement CPR instruction in a nurse faculty. Comput Methods Programs Biomed 2015 Nov;122(2):282-291. [doi: 10.1016/j.cmpb.2015.08.006] [Medline: 26319184]
57. Gaudina M, Zappi V, Bellanti E, Vercelli G. eLaparo4D: A Step Towards a Physical Training Space for Virtual Video Laparoscopic Surgery. In: Proceedings of the Seventh International Conference on Complex, Intelligent, and Software Intensive Systems. 2013 Presented at: IEEE'13; July 3-5, 2013; Taichung, Taiwan p. 611-616. [doi: 10.1109/CISIS.2013.110]
58. Jezek F, Tribula M, Kulhanek T, Matejak M, Privitzer P, Silar J, et al. Surviving Sepsis - a 3D Integrative Educational Simulator Internet. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2015 Presented at: IEEE'15; August 25-29, 2015; Milan, Italy. [doi: 10.1109/embc.2015.7319191]
59. Mattsson SE, Elmqvist H, Otter M. Physical system modeling with Modelica. Control Engineering Practice 1998 Apr;6(4):501-510 [FREE Full text] [doi: 10.1016/S0967-0661(98)00047-1]
60. Tiller M. Introduction to Physical Modeling with Modelica Internet. Switzerland: Springer Science & Business Media; 2012.
61. Fritzson P. Principles of Object-Oriented Modeling and Simulation with Modelica 3.3: A Cyber-Physical Approach. Hoboken, New Jersey: John Wiley & Sons; 2015.
62. Zimmer D. Equation-based modeling with Modelica – principles and future challenges. Simul Notes Eur 2016 Jun;26(2):67-74 [FREE Full text] [doi: 10.11128/sne.26.on.10332]
63. Fritzson P, Engelson V. Modelica - A Unified Object-Oriented Language for System Modeling and Simulation Internet. In: Proceedings of the 12th European Conference on Object-Oriented Programming. 1998 Presented at: ECOOP'98; July 20-24, 1998; Brussels, Belgium p. 67-90. [doi: 10.1007/bfb0054087]
64. Mateják M, Kulhánek T, Šilar J, Privitzer P, Ježek F, Kofránek J. Physiolibrary - Modelica Library for Physiology. In: Proceedings of the 10th International Modelica Conference. 2014 Presented at: Modelica'14; March 10-12, 2014; Lund, Sweden p. 499-505. [doi: 10.3384/ecp14096499]
65. Mateják M. Physiolibrary. URL: http://www.physiolibrary.org [accessed 2019-03-20] [WebCite Cache ID 7712ubnjP]
66. OpenModelica. URL: http://www.openmodelica.org [accessed 2019-03-20] [WebCite Cache ID 771CJRYX6]
67. Dassault Systèmes. DYMOLA Systems Engineering URL: https://www.3ds.com/products-services/catia/products/dymola [accessed 2019-03-20]
68. Kofránek J, Ježek F, Mateják M. Modelica Language - A Promising Tool for Publishing and Sharing Biomedical Models. In: Proceedings of the 1st American Modelica Conference. 2018 Presented at: Modelica'18; October 9-10, 2018; Cambridge, Massachusetts p. 196-205. [doi: 10.3384/ECP18154196]
69. Blochwitz T, Otter M, Akesson J, Arnold M, Clauss C, Elmqvist H, et al. Functional Mockup Interface 2.0: The Standard for Tool Independent Exchange of Simulation Models. In: Proceedings of the 9th International Modelica Conference. 2012 Presented at: Modelica'12; September 3-5, 2012; Munich, Germany p. 173-184. [doi: 10.3384/ecp12076173]
70. Modelica. 2014. Functional Mock-Up Interface for Model Exchange and Co-Simulation URL: https://svn.modelica.org/fmi/branches/public/specifications/v2.0/FMI_for_ModelExchange_and_CoSimulation_v2.0.pdf [accessed 2019-04-27]
71. Flanagan D. JavaScript: The Definitive Guide. Sebastopol , California: O'Reilly Media; 2006.
72. W3Schools. JavaScript Versions URL: https://www.w3schools.com/js/js_versions.asp [accessed 2019-05-06] [WebCite Cache ID 78AEPp3yy]
73. Arseniev A. GrapesJS. URL: https://grapesjs.com/ [accessed 2019-03-20] [WebCite Cache ID 771F7iFCt]
74. Skinner G. CreateJS. URL: https://createjs.com/easeljs [accessed 2019-03-20]
75. Plotly: Graphing Libraries. URL: https://plot.ly/javascript [accessed 2019-03-20]
76. Zakai A. WebAssembly. URL: https://webassembly.org/ [accessed 2019-03-20] [WebCite Cache ID 771Igpmqk]
77. Zakai A. Emscripten. URL: https://emscripten.org/ [accessed 2019-03-20] [WebCite Cache ID 771J7pAPY]
78. Zakai A. Emscripten: An LLVM-to-JavaScript Compiler. In: Proceedings of the ACM International Conference Companion on Object Oriented Programming Systems Languages and Applications Companion. 2011 Presented at: OOPSLA'11; October 22-27, 2011; Portland, Oregon, USA p. 301-312. [doi: 10.1145/2048147.2048224]

79. Falliere N. PNF Software. 2018. Reverse Engineering WebAssembly URL: https://www.pnfsoftware.com/reversing-wasm.pdf [accessed 2019-03-20] [WebCite Cache ID 78IQ6oyZX]
80. Polak D, Šilar J, Ježek F, Mladek A, Kofranek J. Bodylight.js. URL: https://bodylight.physiome.cz/
81. Dassault Systèmes. Code and Model Export URL: https://www.3ds.com/products-services/catia/products/dymola/code-and-model-export/ [accessed 2019-03-20] [WebCite Cache ID 76qcZJQPh]
82. Polák D. GitHub. Bodylight.js-FMU-Compiler URL: https://github.com/creative-connections/Bodylight.js-FMU-Compiler [accessed 2019-03-20] [WebCite Cache ID 771Ltg3vq]
83. Polák D. GitHub. Bodylight.js-OM-Compiler URL: https://github.com/creative-connections/Bodylight.js-OM-Compiler [accessed 2019-03-20] [WebCite Cache ID 771MVY2y9]
84. Adobe. [A New Era of Animation] URL: https://www.adobe.com/cz/products/animate.html [accessed 2019-05-04] [WebCite Cache ID 771N3J5LR]
85. Polák D, Šilar J. BodylightJS. Creating a Simple Project URL: https://bodylight.physiome.cz/docs/simple_project/1_creating_a_simple_project [accessed 2019-04-27]
86. Polák D, Šilar J, Kofránek J. BodylightJS. Pressure volume cardiac simulator URL: https://bodylight.physiome.cz/docs/pvloops/1_pressure_volume_cardiac_simulator [accessed 2019-04-26]
87. Šilar J, Mladek A, David A, Živny J, Kofranek J. Nephron simulator URL: http://physiome.cz/apps/Nephron/
88. Šilar J, Ježek F, Mládek A, Polák D, Kofránek J. Model Visualization for e-Learning, Kidney Simulator for Medical Students. In: Proceedings of the 13th International Modelica Conference. 2019 Presented at: Modelica'19; March 4–6, 2019; Regensburg, Germany URL: http://paperpile.com/b/bfl1h6/jNZY [doi: 10.3384/ecp19157393]
89. David P. Laboratory of Biocybernetics and Computer Learning Support. Simple circulation simulator URL: http://physiome.cz/apps/SimpleCirculation/
90. Kofranek J, David P. Laboratory of Biocybernetics and Computer Learning Support. Cardiac pressure-volume loops simulator URL: https://physiome.cz/apps/pvloops/
91. Ježek F, Kulhánek T, Kalecký K, Kofránek J. Lumped models of the cardiovascular system of various complexity. Biocybern Biomed Eng 2017;37(4):666-678 [FREE Full text] [doi: 10.1016/j.bbe.2017.08.001]
92. Burkhoff D, Tyberg JV. Why does pulmonary venous pressure rise after onset of LV dysfunction: a theoretical analysis. Am J Physiol 1993 Nov;265(5 Pt 2):H1819-H1828. [doi: 10.1152/ajpheart.1993.265.5.H1819] [Medline: 8238596]
93. Sife AS, Lwoga ET, Sanga C. New technologies for teaching and learning: challenges for higher learning institutions in developing countries. Int J Educ Dev 2007;3(2):57-67 [FREE Full text]
94. Geissbuhler A, Bagayoko CO, Ly O. The RAFT network: 5 years of distance continuing medical education and tele-consultations over the internet in French-speaking Africa. Int J Med Inform 2007;76(5-6):351-356. [doi: 10.1016/j.ijmedinf.2007.01.012] [Medline: 17331799]
95. Chavula J, Feamster N, Bagula A, Suleman H. Quantifying the effects of circuitous routes on the latency of intra-Africa internet traffic: a study of research and education networks. In: Nungu A, Pherson B, Sansa-Otim J, editors. e-Infrastructure and e-Services for Developing Countries. Berlin: Springer-Verlag; 2015:64-73.


## Abbreviations

**ADH:** antidiuretic hormone
**e-learning:** electronic learning
**FMI:** Functional Mock-up Interface
**FMU:** Functional Mock-up Unit
**GUI:** graphical user interface
**HTML:** HyperText Markup Language


*Edited by G Eysenbach; submitted 27.03.19; peer-reviewed by M Tiller, S von Mammen; comments to author 18.04.19; revised version received 01.06.19; accepted 18.06.19; published 03.07.19.*

*Please cite as:*
*Šilar J, Polák D, Mládek A, Ježek F, Kurtz TW, DiCarlo SE, Živný J, Kofranek J*
*Development of In-Browser Simulators for Medical Education: Introduction of a Novel Software Toolchain*
*J Med Internet Res 2019;21(7):e14160*
*URL: https://www.jmir.org/2019/7/e14160*
*doi:10.2196/14160*
*PMID:31271154*

Original Paper

# Lumbar Spine Fusion Patients' Use of an Internet Support Group: Mixed Methods Study

Janni Strøm[1,2], RN, MHSn, PhD; Mette Terp Høybye[1,3], MSc, PhD; Malene Laursen[1,4], MD, PhD; Lene Bastrup Jørgensen[1,4], RN, MScN, PhD; Claus Vinther Nielsen[2,5,6], MD, PhD

[1]Elective Surgery Centre, Silkeborg Regional Hospital, Silkeborg, Denmark

[2]Section for Clinical Social Medicine and Rehabilitation, Department of Public Health, Aarhus University, Aarhus, Denmark

[3]Interacting Minds Centre, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

[4]Section for Public Health, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

[5]DEFACTUM, Central Denmark Region, Aarhus, Denmark

[6]Regional Hospital West Jutland, Central Denmark Region, Aarhus, Denmark

**Corresponding Author:**
Janni Strøm, RN, MHSn, PhD
Elective Surgery Centre
Silkeborg Regional Hospital
Falkevej 1-3
Silkeborg, 8800
Denmark
Phone: 45 41177137
Fax: 45 78416299
Email: japete@rm.dk

## Abstract

**Background:** Internet use within health care contexts offers the possibility to provide both health information and peer support. Internet Support Groups (ISGs) for patients may offer advantages, which are not found in face-to-face support. In patients undergoing lumbar spine fusion (LSF), ISGs could have a particular potential, as peer support on the web might bridge the decreased satisfaction with social life and social isolation found within these patients. ISGs might in this way contribute to increasing the functioning and overall health-related quality of life. However, LSF patients may generally belong to a group of citizens not prone to internet and online peer support. However, our knowledge of how LSF patients use ISGs is limited.

**Objective:** The aim of this study was to describe the characteristics of users of an ISG and thematically explore the content of ISG interactions in Danish patients undergoing instrumented LSF because of degenerative spine disorders.

**Methods:** Participants were recruited from a randomized controlled trial and included in a prospective cohort with a mixed methods design. Sociodemographic characteristics and information on psychological well-being (symptoms of anxiety and depression) were obtained at baseline and 1 to 5 weeks before surgery. Usage of the ISG was registered from baseline until 3 months after surgery. All posts and comments were collected, and content analysis was performed.

**Results:** A total of 48 participants comprised the study population, with a mean age of 53 years (range 29-77). Of the participants, 54% (26/48) were female, 85% (41/48) were cohabitating, 69% (33/48) were unemployed, and the majority (69% [33/48]) had secondary education. Approximately one-third of the participants had symptoms of depression (35%, 17/48) and anxiety (29%, 14/48). Overall, 90% (43/48) of the participants accessed the ISG. No correlations were found between sociodemographic characteristics and access to the ISG. Women were more prone to be active users, contributing with posts ($P$=.04). Finally, active users contributing with posts or comments had viewed more pages, whereas passive users, users without posts or comments, had more interactions with the ISG ($P$<.001). The ISG contained 180 conversation threads, generating 354 comments. The 180 conversation threads in the ISG were constituted by 671 independent dialogue sequences. On the basis of those 671 dialogue sequences, 7 thematic categories emerged.

**Conclusions:** Sociodemographic characteristics were not predictors of ISG use in this study, and active use was found to be gender dependent. Content of interactions on the ISG emerged within 7 thematic categories and focused on social recognition, experience of pain or use of pain medication, experience of physical activity or physical rehabilitation, expression of psychosocial well-being, advising on and exploring the ISG, and employment, which seemed to correspond well with the prevalent occurrence of symptoms of anxiety and depression.

***(J Med Internet Res 2019;21(7):e9805)*** doi:10.2196/jmir.9805

**KEYWORDS**

spinal fusion; medical informatics; self-help groups; eHealth; online social networking; social support

## Introduction

### Background

The aim of this study was to describe the characteristics of users of an Internet Support Group (ISG) and thematically explore the content of interactions in an ISG for patients undergoing instrumented lumbar spine fusion (LSF). The ISG was embedded on a platform also containing animated information and training instructions, designed to support LSF patients primarily within the first 3 months after LSF.

Peer support has been found to be applicable in the general population [1], and the desire from patients to connect with each other has been found to grow [2]. During the last decade, the body of literature exploring the use of ISG in cancer patients, patients with depression, patients with HIV/AIDS or other long-term conditions has grown [3-9]. Some of these studies find a positive effect, when providing ISG, on patients' depression, anxiety, or on quality of life [6-9], and it has been suggested that connecting with others on an ISG reduces the sense of isolation [10,11]. Other studies have mixed results or no positive results [12-14]. In addition, some studies even find negative results, as a Web-based survey suggests that the lack of actual physical proximity makes relationships developed within an ISG less meaningful and actually makes patients feel even more isolated [4].

The possibility of providing electronic health (eHealth) interventions is growing with the increasing availability of the internet [15,16-19]. With the use of ISG, peer support is made easier for citizens who, because of their health status, cannot participate in face-to-face groups, and peer support is made easier for citizens in remote areas, citizens with social anxiety, or those who feel uncomfortable disclosing personal experience in a room with others [9,20]. Particularly for patients with anxiety and depression, the ability to remain anonymous might enable them to use ISGs and gain support from peers [21].

In this study, LSF patients were targeted. These patients represent a quite substantial group of back patients, as fusion of the lumbar spine is a commonly performed surgical procedure when treating various conditions of the spine [22]. Approximately 488,000 spinal fusions were performed during US hospital stays in 2011 (3.1% of all operating room procedures) [23], lumbar fusions being the most common type of fusion performed, approximately 210,000 operations in the United States each year [22].

Introducing ISGs to patients undergoing LSF might have potential, as anxious and depressed patients are found to be more prone to take part in an ISG [24], and as symptoms of anxiety and depression are found in approximately one-third of these patients [25-28]. Furthermore, limiting seated transportation is recommended during the first 6 weeks after surgery, and active rehabilitation starts no sooner than 3 months after surgery [29]. This limits LSF patients' ability to attend face-to-face meetings in this period, and studies find a significant decline in satisfaction with social life after LSF [30], predicting worse score in health-related quality of life after surgery [31]. Thus, several LSF patients are found to have symptoms of anxiety and depression and to be bound to their homes, and thus they might be more prone to use an ISG. In a previously published randomized controlled trial (RCT), we examined the effect of the total Web-based platform (including animated information and training instructions, a click-through diary, written information, and the ISG) on symptoms of anxiety and depression and on pain. We found no additional effect of the Web-based platform on any of the outcome parameters [14].

### Objective

We need to further explore the use of an ISG within this group of patients. Higher age and lower socioeconomic background seem to characterize people who are less frequent internet users and who potentially benefit less from Web-based support [6,32-35], characteristics which generally match those of LSF patients [36]. Thus, we need to explore and describe the characteristics of users and describe the content of the Web-based interaction in an ISG provided to LSF patients, to inform future design and use of such interventions to this patient population. In addition, most studies exploring the use of ISG use already established forums, and little is known about newly initiated forums [37].

## Methods

### Study Design and Setting

This study's population comprises the patients enrolled and randomized from September 2015 to May 2017 [14] to an intervention group in a 2-arm RCT at a single-center orthopedic spine department in Denmark.

The population comprises 48 participants receiving access to an ISG as part of their pre and postsurgical care. All patients were scheduled for first time 1- to 3-level instrumented LSF because of degenerative disorders. Patients were excluded if they were below the age of 18, if they had a known psychotic disorder, if they were unable to communicate in Danish, or had no access to the internet. Patients were invited to participate in the study by a study nurse, receiving written and verbal information, when the patients attended the outpatient clinic and were scheduled for surgery.

Final decision of participation was made, and written consent was collected at a following baseline visit 1 to 5 weeks before surgery. At baseline, participants were introduced and received access to the ISG, and a 15-min hands-on introduction in the use of the website was provided by a study nurse. During the introduction, all participants were encouraged to share their experience, thoughts, and questions on the ISG, and all were encouraged to use a respectful tone. Participants were told to use the ISG in a way that made sense to them and fulfilled their needs. During the study, the participants received no notifications of activity on the ISG. Posts or comments would only be visible when the participants choose to access the ISG.

Patients were continuously included as they attended the outpatient clinic, but because of a variation of patient flow, the speed of inclusion was uncertain. Thus, 6 former patients were invited as facilitators to start the dialogue and post updates in the common space, creating activity and engaging the first participants entering the group. These 6 facilitators were not excluded from the ISG during the study period, and 1 of them contributed to the dialogues throughout the study. We did not include these former patients in the descriptive analysis, describing the characteristics of users during the first 3 months after surgery. The main reason for this was that they had a different role than those participating in the study, as they were committed differently. However, their comments and posts are included in the content analysis, as these were part of the conversational context of what was posted by study participants. The content data were collected from the first participant through to 6-month follow-up of the last attending participant.

It was decided to promote implementation by providing participants with easy access. This was decided as a previous study introducing a similar Web-based platform for patients undergoing total hip replacement found the access rate as low as 61% [38]. With reference to theories within implementation [39], attention was given toward easy access. Thus, patients who were not in possession of a tablet were offered the use of one from baseline and until 3 months after surgery.

The science ethics committee was notified of the study, and it did not find that permission was required (J.no. 1-10-72-36-15). Data management and security were approved by the Danish Data protection agency (J.no. 2014-41-3583). In line with the Helsinki Declaration [40], patients were informed about the study both in writing and verbally and had at least 24 hours to consider their participation.

### The Internet Support Group

The ISG could be accessed from any browser through a designated website. The ISG was closed to the public, and participants logged on using an individual password. Participants were assured that all data were kept for research purposes only and that all correspondence was kept in accordance with data management and security regulations. Furthermore, all participants were informed that it would not be possible to identify any of the participants in any published work. A researcher and a study nurse provided technical support if this was needed, and the researcher would intervene if any offensive remarks were posted or if an aggressive tone was used. No technical support or mediation of such behavior was ever needed, and no intervention or moderation was provided by either the researcher or the study nurse.

The ISG comprised a message board visible to all participants (Figures 1 and 2). In this space, participants could post their experience, thoughts, or questions for other participants to comment on. On a designated page, each participant could upload a picture, note date of birth and date of surgery, and write a personal or background story. No restrictions were made on as to what this personal or background story should include, and no restrictions were made on the length of the post or how frequently participants could write in this space (Figure 3). Everything written here was visible for all participants.

**Figure 1.** The message board visible to all lumbar spine fusion patients assigned to an Internet Support Group. The board is a reconstruction, names are invented, pictures are stock photos, and the text is fictional, based on inspiration from real posts.

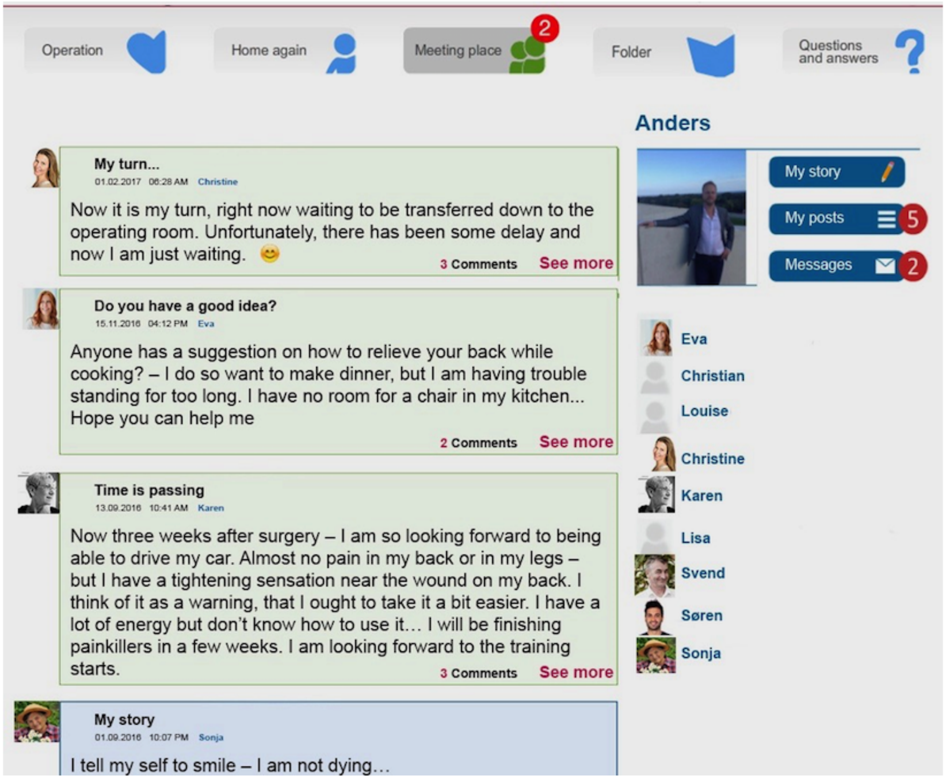



**Figure 2.** The message board with comments, visible to all lumbar spine fusion patients assigned to an Internet Support Group. This message board is a reconstruction, names are invented, pictures are stock photos, and the text is fictional, based on inspiration from real posts.

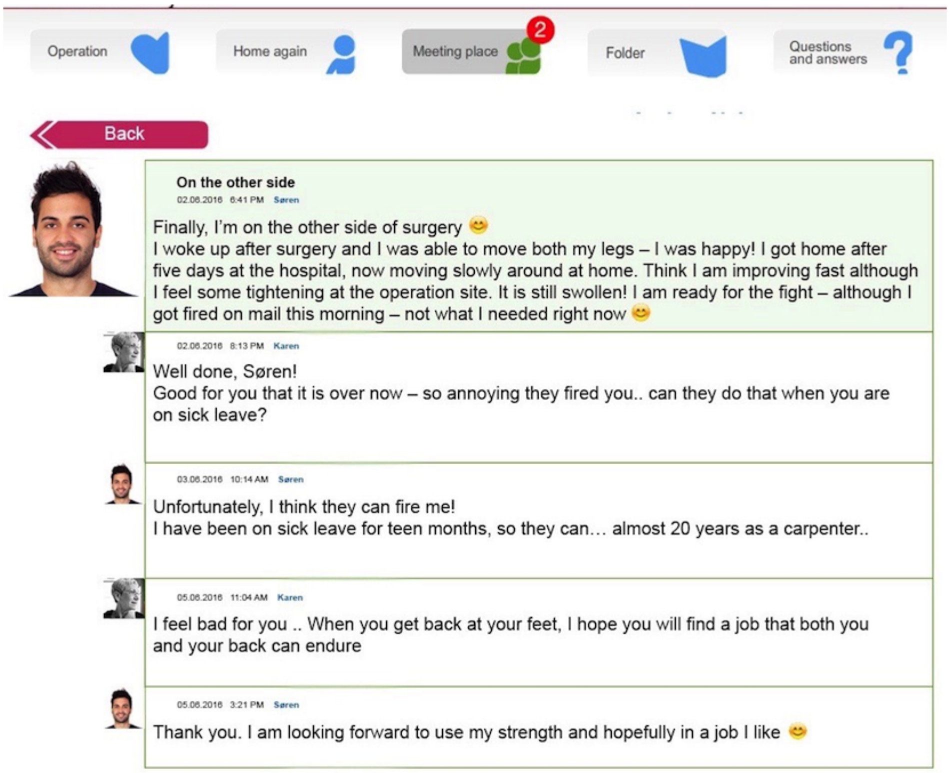
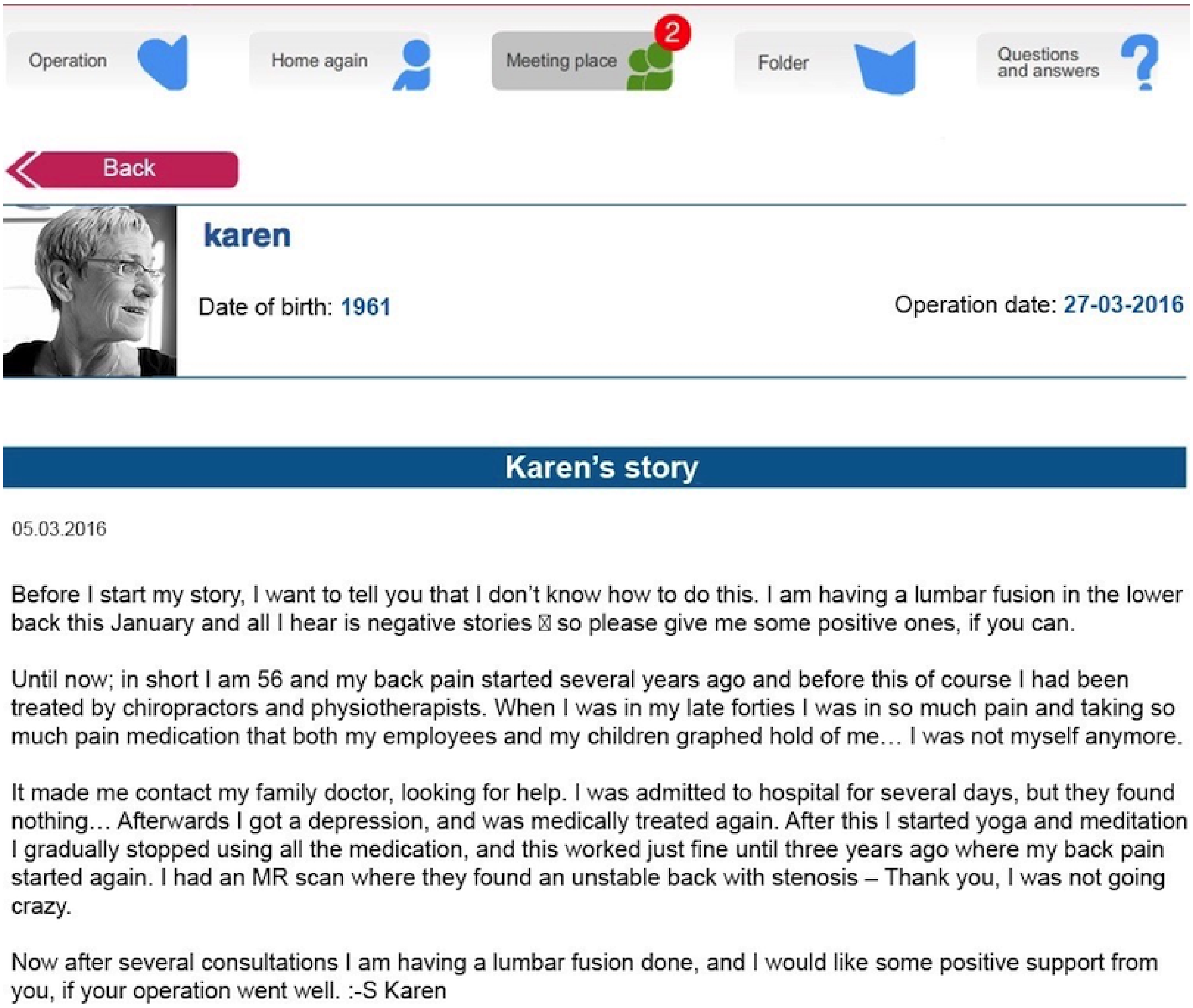


**Figure 3.** Personal page on the Internet Support Group platform, with the possibility to upload a picture, note date of birth and surgery, and write a personal background story. The page is a reconstruction, the name and dates are invented, picture is a stock photo, and the text is fictional based on inspiration from a real story.




### Data Collection

Data collection to describe the characteristics of the use of the ISG during the first 3 months after surgery was partly done in person by a study nurse at baseline visit and partly by manually tracking activity on the ISG until the participant had passed 3 months after surgery. Baseline data were collected by a study nurse, and the data comprised gender, age, sociodemographic background, and psychological well-being. Marital status was classified as married/cohabiting or living alone (including widow, single, or divorced). Educational status was classified into 3 categories, using the International Standard Classification of Education 2011 [41]: basic education (early childhood education, primary education, and lower secondary education), secondary education (upper secondary education), and higher education (postsecondary nontertiary education, short-cycle tertiary education, bachelor's or equivalent, master's or equivalent, and doctoral or equivalent level). Employment status was classified as 1 of 3 categories: (1) employed/full or part-time, (2) pensioner/other (includes participants not employed for other reasons than illness or unemployment, such as housewife, on leave, or student), and (3) sick leave/unemployed. Psychological well-being was obtained using self-administered, paper and pencil Hospital Anxiety and Depression Scale (HADS), which was distributed in person. HADS is a 14-item scale, in 2 subscales for anxiety and depression. The cut-off point has been identified at a score of 8/21 for symptoms of anxiety and depression [42]. For anxiety, this gave a specificity of 0.78 and a sensitivity of 0.9. For depression, this gave a specificity of 0.79 and a sensitivity of 0.83 [42].

### Data Collection to Analyze the Content of Interactions

From baseline until 3 months after surgery, participants' activity on the ISG was monitored manually. This was done by tracking the use of google analytics, which is a free Web analytics tool generating detailed statistics concerning user behavior on websites. Google analytics were used, including location data, browser data, device type, event type, and event time; user-generated content data, including all posts, comments and stories by individual users, and personal data, including location, access date, date of operation, and from which device the participant gained access.

Activity was measured as interactions; 1 interaction is when 1 participant interacts with the ISG. An interaction often comprised a group of page views taking place within the same session. Distinctions were made between posts and comments. A post is a new upload of a question, update, or other, and a comment is an answer or comments written in a thread of an uploaded post.

Participants accessing the ISG without making posts or comments are defined as passive users, and those participants who contribute with posts or comments are defined as active users. The entire group entering the ISG will be referred to as users. Time spent on pages was not utilized, as it was not possible to ascertain whether the ISG was used when the pages were open.

### Statistical Analysis

All data were coded to compute the data statistically and then entered twice into Excel. The data were then transferred to StataCorp. 2017. STATA Statistical Software: Release 15, here all statistical analyses were performed.

Demographics were utilized to describe the sample. Descriptive statistics were done using frequencies and percentages to describe the sample profile and summarize data. Means and standard deviations were reported for continuous variables. Nonparametric data were analyzed using Spearman correlation tests to detect correlation among variables. Kruskal-Wallis ranks tests and Wilcoxon rank-sum test were used to establish any significant differences among unordered groups.

### Content Analysis

Qualitative data, comprising posts and comments from the ISG in a 2-year period (from first patient entered until the last patient attended 6-month follow-up), were collected and analyzed to explore the content of interactions on the ISG. All posts and comments were collected manually from the ISG and managed using NVivo qualitative data analysis software; QSR International Pty Ltd. Version 12, 2018. Data were analyzed using a qualitative content analysis [43]. Taking an overall inductive approach, all correspondence was read through to get an impression of the general context. It was then organized using cross-sectional indexing [44], on the basis of similarities and variations in the dialogue among participants, to produce an overview of the thematic indicators and analytical categories of use of the ISG. On the basis of these inductive categories, an explanatory synthesis was aggregated from the categories to condense the impression of interactions, presented as individual thematic categories below.

## Results

### The Characteristics of Internet Support Group usage

This paper represents 48 patients undergoing instrumented LSF, and out of these participants, 57% of the participants were females. The mean age was 53 years (range 29-77). A total of 69% of the participants completed secondary education, whereas 27% of the participants had completed basic education, and only 6% of the participants had completed higher education. Preoperatively, a total of 40% of the participants were on sick leave or for other reasons unemployed; 31% of the participants were employed, and 29% of the participants were pensioners/other. Thus, approximately 70% of the participants were outside the labor market. A clear majority of participants were married/cohabitating (84%; Table 1).

**Table 1.** Sociodemographic characteristics of the participants (n=48).

| Characteristics | Values |
|---|---|
| **Gender, n (%)** | |
| Male | 22 (46) |
| Female | 26 (54) |
| Age (years), mean age (range) | 53 (29-77) |
| **Marital status, n (%)** | |
| Married/cohabitating | 41 (84) |
| Living alone[a] | 7 (15) |
| **Education, n (%)** | |
| Basic education[b] | 12 (25) |
| Secondary education[c] | 33 (69) |
| Higher education[d] | 3 (6) |
| **Employment status, n (%)** | |
| Employed full/part time | 15 (31) |
| Pension/other[e] | 14 (29) |
| Sick leave/unemployed | 19 (40) |

[a]Includes participants who are widowed, single, or divorced.

[b]Basic education level comprised International Standard Classification of Education levels 0-2.

[c]Secondary education comprised International Standard Classification of Education levels 3.

[d]Higher education level comprised International Standard Classification of Education levels 4-8 [41].

[e]Includes participants not employed for reasons other than illness or unemployment, such as housewives, those on leave, or students.

A total of 59% of the participants (n=29) chose to borrow a tablet. No correlations were found among gender, age, or sociodemographic data or among any of the valuables of use in relation to whether the participant chose to borrow a tablet or not (*P*>.45).

The 48 participants had a total of 933 interactions on the ISG during their first 3 months after surgery. A total of 90% of the participants (n=43) were users of the ISG, and 5 participants (10%) never used the ISG. The mean number of interactions for the 48 participants was 19.4 (range 0-90, SD 19.28); however, the 3 most active users had a total of 25.5% (238) of all interactions.

In the 933 interactions, the total number of page views was 2093. The mean number of page views for each of the 48 participants was 42.7 (range 0-312, SD 62.84). Of the 48 participants, a total of 48% participants (n=23) participated with posts or comments. The total number of posts and comments the during first 3 months after surgery was 288, and the mean number of posts or comments for the 48 participants was 6 (range 0-61, SD 13). Spearman rho was performed, looking for correlation between variables of use of the ISG and age and education. No significant correlations were found (Table 2). Analysis was performed to detect differences between groups (Tables 3 and 4). No differences were found except for gender, indicating that contributing posts were more common in women.

At baseline, 14 participants (29%) scored 8 or more on the HADS anxiety subscale, indicating the presence of symptoms of anxiety, and 17 participants (35%) scored 8 or more on the HADS depression subscale, indicating the presence of symptoms of depression. A total of 8 (16%) of the participants had symptoms of both depression and anxiety, leaving only 25 participants (52%) without either symptoms of anxiety or depression (Table 5).

No significant differences were found among groups regarding the use of the ISG and the presence or absence of anxiety and depression (Table 6). However, participants with symptoms of anxiety tended to be more prone to contribute with posts or comments on the ISG than those without anxiety.

**Table 2.** Correlation between variables of use and demographic data. No significant correlations were found between age or education and the variables of Internet Support Group use (Spearman correlation).

| Demographic data | Interactions | | Page views | | Posts | | Comments | |
|---|---|---|---|---|---|---|---|---|
| | Spearman rho | *P* value | Spearman rho | *P* value | Spearman rho | *P* value | Spearman rho | *P* value |
| Age | –0.0706 | .63 | 0.0530 | .72 | –0.1305 | .37 | –0.1110 | .45 |
| Education | –0.0920 | .53 | –0.1702 | .24 | 0.0220 | .88 | –0.1076 | .46 |

**Table 3.** Correlation between activity on the Internet Support Group and the participants' employment status. No significant correlations were found between employment status and the variables of Internet Support Group use (Kruskal-Wallis test).

| Group | Rank sum | *P* value |
|---|---|---|
| **Interaction (n)** | | **.90** |
| Employed/full or part time (15) | 88.00 | |
| Pension/other (14) | 329.50 | |
| Sick leave/unemployed (19) | 507.50 | |
| **Page views (n)** | | **.80** |
| Employed/full or part time (15) | 388.00 | |
| Pension/other (14) | 319.50 | |
| Sick leave/unemployed (19) | 517.30 | |
| **Posts (n)** | | **.82** |
| Employed/full or part time (15) | 399.00 | |
| Pension/other (14) | 344.00 | |
| Sick leave/unemployed (19) | 481.00 | |
| **Comments (n)** | | **.95** |
| Employed/full or part time (15) | 370.00 | |
| Pension/other (14) | 362.50 | |
| Sick leave/unemployed (19) | 492.00 | |

**Table 4.** Correlation between activity on the Internet Support Group and marital status and gender (n=48).

| Group | n (%) | Interaction | | Page views | | Posts | | Comments | |
|---|---|---|---|---|---|---|---|---|---|
| | | Rank sum | *P* value | Rank sum | *P* value | Rank sum | *P* value | Rank sum | *P* value |
| **Sex** | | | **.31** | | **.35** | | **.04**[a] | | **.10** |
| Male | 22 (46) | 475.5 | | 478.5 | | 442.5 | | 451 | |
| Female | 26 (54) | 749.5 | | 746.5 | | 782.5 | | 774 | |
| **Marital status** | | | **.85** | | **.87** | | **.71** | | **.26** |
| Married/cohabitating | 41 (85) | 1032 | | 1030 | | 1036.5 | | 1062.5 | |
| Living alone | 7 (15) | 193 | | 194 | | 188.5 | | 162.5 | |

[a]Significant correlation was found between female gender and the contribution of posts on the Internet Support Group (Wilcoxon Rank-sum).

**Table 5.** The presence of anxiety and depression at baseline (n=48).

| Anxiety/Depression | +Depression, n (%) | –Depression, n (%) | Total, n (%) |
|---|---|---|---|
| +Anxiety | 8 (17) | 6 (14) | 14 (29) |
| –Anxiety | 9 (19) | 25 (52) | 34 (73) |
| Total | 17 (35) | 31 (65) | 48 (100) |

**Table 6.** Use of Internet Support Group (ISG) in participants with or without anxiety and depression at baseline (n=48). No significant correlations were found between the presence of anxiety or depression and variables of ISG use (Kruskal-Wallis ranks test).

| Group | n (%) | Rank sum | *P* value |
|---|---|---|---|
| **Interactions** | | | |
| **Anxiety** | | | **.85** |
| – | 34 (73) | 841.50 | |
| + | 14 (29) | 383.50 | |
| **Depression** | | | **.35** |
| – | 31 (65) | 868.50 | |
| + | 17 (35) | 356.50 | |
| **Page views** | | | |
| **Anxiety** | | | **.67** |
| – | 34 (73) | 830.50 | |
| + | 14 (29) | 394.50 | |
| **Depression** | | | **.17** |
| – | 31 (65) | 888.00 | |
| + | 17 (35) | 337.00 | |
| **Posts and comments** | | | |
| **Anxiety** | | | **.07** |
| – | 34 (73) | 773.50 | |
| + | 14 (29) | 451.50 | |
| **Depression** | | | **.59** |
| – | 31 (65) | 848.00 | |
| + | 17 (35) | 377.00 | |

**Table 7.** Interactions and page views in groups who are passive users or active users (n=48).

| Group | n (%) | Rank sum | *P* value |
|---|---|---|---|
| **Interactions** | | | **<.001**[a] |
| Passive users | 25 (52) | 485.00 | |
| Active users | 23 (48) | 374.00 | |
| **Page views** | | | **<.001**[b] |
| Passive users | 25 (52) | 463.50 | |
| Active users | 23 (48) | 761.50 | |

[a]Significant correlation was found between passive users and interactions.

[b]Significant correlation was found between active users and page views.

Table 7 shows the analysis of user variables comparing passive users with active users. Those users who were active had the most page views and passive users had the most interactions (Table 7).

## Interaction on the Internet Support Group

The total number of interactions on the ISG was 3357 in a 2-year period from the first participant entered until the last included participant reached 6 months follow-up after surgery. The ISG contained 180 conversation threads, generating 354 comments. The 180 conversation threads in the ISG were constituted by 671 independent dialogue sequences. On the basis of those 671 dialogue sequences, 7 thematic categories emerged: social recognition, experience of pain or use of pain medication, experience of physical activity or physical rehabilitation, expression of psychosocial well-being, and advising on and exploring the ISG and employment. Examples of findings within the 7 categories are shown in Textbox 1, and a short description of each of the categories is provided below. The categories are presented starting with the one including the largest number of threads and dialogue sequences and ending with the one with the fewest.

**Textbox 1.** The 7 categories and their associated dialogue sequences extracted from dialogs among the 23 active users of the Internet Support Group, including the 6 former patients who had a facilitating role.

Social recognition

- Dialogue sequence 3: "I would like to hear more from you, lets hook up on Facebook."
- Dialogue sequence 6: "Now it is holiday for a lot of people...I hope for sun to shine."
- Dialogue sequence 170: "I was wondering, where are you all from?"
- Dialogue sequence 175: "I would like to meet up with you, I will text you."
- Dialogue sequence 209: "Today I have been visiting Eva, it was very nice. I am grateful that she wanted to meet up with me and share her experiences."
- Dialogue sequence 120: "Hello and welcome to all new.."
- Dialogue sequence 303: "The best to you all...and have a nice weekend."

Experience of pain or use of pain medication

- Dialogue sequence 308: "I have had a lot of pain since the operation...I guess that is what to expect."
- Dialogue sequence 317: "I am in a lot of pain, but until now I have been able to keep it on a bearable level using medication and red wine at night. That doesn´t work anymore...now I can´t walk."
- Dialogue sequence 332: "I decreased the medication to quickly...I have no patience."
- Dialogue sequence 389: "...I do not use morphine anymore and my nausea is gone as well. I take paracetamol four times a day and that is enough."

Experience of physical activity or rehabilitation

- Dialogue sequence 426: "Hello! Today I have been walking for an hour, but it was hard, I had to rest when I got home. We were walking quite calm, but I had to rest during the walk as well."
- Dialogue sequence 454: "I am only able to do the exercises some days and I can´t do all of the exercises. I just started walking outside and it is eight weeks since I was operated."
- Dialogue sequence 483: "Yes, you need to be careful and not pressure you self too much. I walked a bit more than one kilometre yesterday, it went okay. But when I came home my back hurt...I had gone too far."
- Dialogue sequence 503: "I start rehabilitation next week, I am looking forward to it."

Expression of psychosocial well-being

- Dialogue sequence 521: "I wept of pain and experienced the feeling of being small and miserable."
- Dialogue sequence 524: "I do believe it will get better, I have always said that I would turn 107."
- Dialogue sequence 531: "I want to hear how you are...I am going out of my mind"
- Dialogue sequence 534: "It is very difficult when I am used to being social and active, it almost makes me cry. I am almost eight weeks after surgery and now I think it is time it got better. I need more patience – I am looking forward to hearing from you all."

Expression of everyday activities

- Dialogue sequence 565: "I have found a new rhythm for my day, I watch TV, I do crosswords and I relax, it is nice."
- Dialogue sequence 575: "Can anyone suggest things I should prepare for when I get home."
- Dialogue sequence 596: "to those of you who are past the operation...is it possible to bend forward to tie your shoes? Or how do you do?"
- Dialogue sequence 600: "any suggestions of how-to relief the back when cooking? I want to cook; however, I can´t stand by the kitchen table for so long."

Advising on and exploring the Internet Support Group

- Dialogue sequence 608: "I have worked out how to get in (on the ISG) using my PC, so I am still here."
- Dialogue sequence 618: "God morning everyone – I was thinking how do we get more activity here? I know that when all newly operated may not have the energy. But the rest of us? Any ideas?"
- Dialogue sequence 621: "I hope you all want to join; our experiences are worth gold for others."
- Dialogue sequence 625: "Just write something about the weather, if you don´t want to share your health – that is okay. It is better to write something than nothing."

Employment

- Dialogue sequence 634: "I do some voluntary work at a reception, I hope to start again this January."
- Dialogue sequence 643: "I am now 3 months and 15 weeks and I just started working 4 hours each week...Yes, it is not much, but it is very important for me."
- Dialogue sequence 647: "Nice to hear that you found a rhythm working, which suits your back – It is a fine line, which is hard to find."
- Dialogue sequence 655: "When you get better I hope you will find a job than both you and your back can manage."

### Social Recognition

The category represented the largest number of threads (n=145, 81%) and largest number of dialogue sequences (n=307, 46%) related to social recognition in and of the ISG. This category had 3 subthemes: The largest subtheme comprising a substantial number of the total dialogue sequences (n=233, 35%) was supportive comments, where group members welcomed new members to the ISG, expressed thinking of each other, and expressed concern and well-wishes for other group members. Another subtheme within social recognition was the invitation to other group members to interact outside the ISG. Finally, a number of dialogue sequences comprised more ordinary chitchat unrelated to the LSF surgery and recovery; they also comprised exchanging stories on how to spend their vacation, enjoying the spring or about serious subjects, such as losing a family member.

### Experience of Pain or Use of Pain Medication

The discussion of pain or pain medication was prevalent in 56 threads (31%), comprising 116 dialogue sequences (17%). Dialogues concerning pain medication either contained posts or comments on how to decrease the use of pain medication or expressions of what was found to be the side effect of the medication. Finally, expressions of either increase or decrease of pain after surgery was a recurrent topic of conversation.

### Experience of Physical Activity or Physical Rehabilitation

In 45 threads (25%) and 82 dialogue sequences (12%), posts or comments focused on the experience of walking or seeking to motivate each other to walk. Finding the balance between walking and resting was another recurrent line of conversation, as well as the start of active rehabilitation.

### Expression of Psychosocial Well-Being

Psychosocial well-being was a key focus in 34 threads (19%) of the discussion, appearing in 54 dialogue sequences (8%). These could be divided in 2 subthemes, 1 which was positive statements of future expectations, such as expressions of being optimistic of future recovery. The other was expressions of anxiety, lack of energy, or hard times.

### Expression of Everyday Activities

The request of information from peers was a key theme in the use of the ISG. A total of 24 threads (13%) contained 42 dialogue sequences (6%) concerning request of information. A total of 35 of these focused on expression of everyday situations, sharing with the groups their experience on how to manage cooking, pass time, etc. A total of 9 dialogue sequences focused on health-related information, such as how to tend to the bandages or the itching of the wound.

#### Advising on and Exploring the Internet Support Group

In 18 threads (10%), 28 dialogue sequences (4%) were passing advice among members on the use of the ISG. These dialogue sequences were a mixture of advice from one to the other of where to write what type of content, invitations to upload posts and to share stories, and dialogues on how to create more activity in the ISG.

### Employment

A common concern for many of the LSF patients who used the ISG was related to employment. This came up in 16 threads (9%), comprising 42 dialogue sequences (6%), for example, commenting on the number of hours of employment it was possible to undertake, the lack of employment, or the experience of being fired following the LSF surgery.

## Discussion

### Socially Constructed Patterns in Internet Support Group Usage

Gender distribution was almost equal in this study, with no significant differences between men and women with regard to the use of the ISG, which is in line with previous findings [35,45,46]. However, significantly, more women were active users, contributing by uploading posts. This is in line with other findings where women were found to be more than 4 times as likely to be active users, contributing with written uploads [47]. Explanations are offered for this discrepancy, focusing on socially-constructed patterns of behavior [46]. The general division of labor and the pay gap within and between occupations are generally responsible for larger incomes among men. Thus, it rationally makes more financial sense that the woman in a partnership takes time off to perform care duties, making them more acquainted with and prone to engage in health care information on the internet [46]. Another explanation proposed as to why men may be averse to asking questions and posting comments in an ISG is that they generally act according to a set of masculine values: asking questions requires a confession of ignorance or need, which may pose a threat to masculinity [46]. Furthermore, men generally tend to consider health situations to be less harmful, generating less motivation to ask questions and seek information from others [46]. Supporting this argument, the Comprehensive Model of Information Seeking points out that a person's needs or perceptions of risk influence the degree to which the ability to do something about a health problem is considered to be realistic, thereby generating information seeking behavior [48]. That perceived risk influences social behavior is also found in another study published from the United Kingdom that included 863 depressed and anxious participants [24]. They found that anxious and depressed participants were more prone to take part in an ISG [24]. Participants with symptoms of anxiety in this study also tended toward being more prone to uploading posts and comments. Thus, the degree of contribution in this study seems to be gender related and could further be related to mental health state.

There was no sociodemographic variable in this study that correlated with the use of the ISG. Lack of such predictors was also found in the previous mentioned study from the United Kingdom [24]. However, others find that socioeconomic status, comprising education and employment status, correlates with ISG use [6,35,47,49]. Some of the explanation for the discrepancy among findings might be found in the results of a US study, with data drawn from 2358 participants [47]. In that study, a differentiation was made between seeking information and engaging in social media. It was concluded that lower socioeconomic status, older age, and male gender were associated with less likelihood to engage in eHealth activities, such as seeking information; it was also concluded that lower socioeconomic status was associated with a greater likelihood to use health-related social media [47]. In the context of this study, the ISG was hosted on a website with additional animated information. Thus, the website incorporated both information and social interaction, and it might appeal to a broader audience, limiting the possibility of finding predictors for ISG use.

### Provision of Information and the Nature of Posts

Provision of animated information on the website might have had further influence on the ISG use. We chose not to extract dialogue sequences categorized as seeking of information, as the request of information or giving recommendations did not dominant the correspondence. Posts and comments were to a larger extent characterized by the expression of sharing experience, supporting each other and engaging in common dialogues. In a recently published study, done on bariatric surgery support groups and pages on Facebook, 11% of the content indicated seeking of information and 53% of the content contained the provision of information and recommendations [50]. The additional animated information might have reduced the need for information from peers.

Furthermore, post hoc analysis found no posts or comments containing incorrect or contradictive information in this study. Using ISGs, patients are found to be concerned about the quality of information being shared, and they occasionally encounter contradictive information, actually adding to patient's insecurity rather than the opposite [5]. The animated information provided on the website, together with the ISG, may have reduced some of these misconceptions. Thus, the provision of an ISG on a website, together with information, might appeal to a larger audience and might also reduce the sharing of misinformation. In addition, the ISG was used in the context of a research study, where the participants were being monitored, which also might have made them reluctant to exchange information.

### Social Recognition

Social recognition was the largest thematic category emerging from the posts and comments, with a substantial number of dialogue sequences (n=307, 46%), and with supportive posts and comments being the largest subtheme with 233 (35%) dialogue sequences. A possible explanation for the many supportive comments and posts might be found in the characteristics of the patient group. There seems to be a distinction between which condition the participant has and the nature of posts and comments made. Mental health disease ISGs comprise more emotional support, although it seems that the requests for information relate more toward physical conditions [51]. LSF is performed because of a physical condition; however, as presented above, half of the participants had symptoms of anxiety or depression, and 16% had symptoms of both, which might contribute to the large amount of supportive posts. Furthermore, the 7 thematic categories emerging from the content analysis, social recognition, experience of pain or use of pain medication, experience of physical activity or physical rehabilitation, expression of psychosocial well-being, advising on and exploring the ISG and employment, correspond to a large extent to the factors, which have been found to be associated with symptoms of anxiety and depression within patients undergoing spinal surgery, such as pain, information, disability, employment, and mental health [52]. Thus, the often occurrence and nature of supportive posts might be because of the prevalent occurrence of symptoms of anxiety or depression, which have been found to accompany one-third of spine surgery patients [25-28].

### Implementation of the Internet Support Group

A greater percentage of participants accessed the ISG in this study than in other studies evaluating a wider use of ISGs and in participants with other diagnosis [24,35]. There may be several reasons for this degree of participation. A total of 3 possible explanations will be explored below. First, a choice was made to increase accessibility, and patients were offered the use of a tablet. This choice may have increased the participants' perceived ease of use and accessibility of the ISG, which may have facilitated a more positive attitude and frequency of use. That ease of use influences acceptance and adoption is in concordance with the theory of the Technology Acceptance Model [53] and the previously mentioned implementation theory [39]. Second, all participants in this study agreed to participate, knowing that they would have to relate to an ISG. Agreeing to these terms may have created more motivated users. A third possible explanation could be related to the ISG being embedded on a website, with an additional number of separate components available as well. A broader assortment of support may be appealing and useful for a wide range of participants, attracting both those seeking information and those seeking health-related social media. The influence of perceived usefulness is in accordance with the Technology Acceptance Model and implementation theories, and it is found to support a positive attitude and a behavioral intention toward use of a new technology [39,53]. Thus, in the light of the above the choice to increase accessibility, the motivation and the perceived usefulness might have been facilitators, increasing the use of the ISG in this study.

A small number of the participants in this study were responsible for the majority of interactions. The 3 most active users contributed 25% of the total number of interactions. This phenomenon that a few contribute a lot is known as the peer leader phenomenon, characterized by a high posting frequency.

This tendency is found to the extent that 1% of participants are seen to contribute 75% of all posts [54]. Such peer leaders identify themselves as active help providers, tending to provide a high level of social support, which contributes to an increased effect of the ISG if social support underpins the improvement [54]. There is a high amount of supportive comments comprising 35% of all extracted dialogue sequences in this study. This provides some indication of the nature of the social behavior in the ISG employed in this study, as there seems to be a positive tone among the participants.

Approximately half of the participants in this study (48%) contributed by posting or making comments. It is known that members of online groups often begin their membership as visitors or so-called lurkers, who just observe, and that not all members shift from being visitors to active users [55]. Some retain a passive behavior, as members are mainly interested in the information they can access through the online group [55]. In this study, the passive users are those with the most page views, which may indicate that they were interested in the information they could retrieve. It is possible that the fact that interactions within ISGs comprise mainly of text has an impact on patterns of use to the extent that only half of the users were active and contributors in this study. This type of Web-based interaction does require that participants are comfortable writing and reading information on the Web. eHealth literacy (defined as the ability to seek, find, understand, and appraise health information from electronic sources and apply knowledge gained to address or solve a health problem) [56] is known to correlate with social position, older age, and chronicity of disease [47], and it may be one of the reasons why some of the participants in this study did not contribute actively to the ISG. It is suggested that education might be a more salient proxy for ISG use than income, indicating that determinants of eHealth literacy may be important predictors [47]. It was not possible to find correlations in this study confirming the level of education as a predictive factor for ISG use, the reason for this might be found in the high percentage of participants (69%) with a secondary education. However, it seems plausible that eHealth literacy must be taken into consideration when developing internet interventions. Failure to do so could increase the already existing inequality of health care. Further research should be done accessing eHealth literacy and its influence on eHealth engagement across social groups.

In the literature, expressed limitations of ISG use can also be found. Even though it is acknowledged that the key reason for citizens to participate in an ISG is the connection with others, a Web-based survey suggests that the lack of actual physical proximity makes relationships developed within an ISG less meaningful and actually makes patients feel even more isolated outside of the ISG [5]. Looking at the content of posts and comments, a subtheme did reveal invitation to other group members to interact outside the ISG.

## Limitations

This study has several imitations. First of all, the use of google analytics, together with user data and personal data, to uncover activity on the ISG has limitations. The quality of data could have been increased by the use of a unique user ID added to the platform's user accounts. This could have automatically identified behavior on the ISG, providing more accurate data and a clearer picture of events than that provided in our manually generated process.

The number of participants is small, limiting the strength of the analysis and making it difficult to draw statistically sound conclusions. However, our sample does not deviate markedly from those in the Danish population who report having a spinal condition in relation to gender, age, educational level, and employment status [33]. Moreover, the prevalence of both anxiety and depression in this sample of LSF patients is equal to that found in the literature among a similar group of patients [25,27,28,57]. However, further research with a larger group of participants should be done, and it might uncover further knowledge on the characteristics of use of an ISG in patients undergoing LSF.

All participants not owning a tablet were offered to borrow one, and thus half of the participants (59%) chose to borrow such a device, clearly influencing the use of the ISG and the generalizability of the study results, as we are not able to conclude on the use of the ISG in a home without the use of a tablet.

All the participants did not access the ISG at the same time; the recruitment process was consecutive and slow in periods. It is not possible to know what influence this had on the frequency of use. To accommodate the slow recruitment process, 6 former patients were invited to help engage participants. The role of these 6 patients is not further uncovered within this study; however, their role as moderators may have had an influence on the usage, the content of uploaded posts and comments, and the quality of the study. Furthermore, all participants were aware that their interactions were being studied, and thus some might be more reluctant to engage in the ISG; this could be uncovered in future studies.

Activity on the ISG is seen as a positive contribution; however, the perceived value of the different contributions was not considered. We did look at the content of posts and comments that contributed to the knowledge of behavior within an ISG; however, an exploration of how the posts and comments were received could have been relevant and would provide additional knowledge on behavior in an ISG.

The full potential of internet use within health care has not yet been reached, for example, with regard to dissemination of information, establishing health community networks, blogging, and establishing support groups. It is important to acknowledge that this global availability of information and support comes with difficulties. We need to strive to accommodate these difficulties by providing high-quality eHealth accessible to all.

## Conclusions

This paper contributes to the literature on the use of ISGs within health care and especially within the group of patients undergoing LSF. Socioeconomic status was not an important barrier in this study, and it was not possible to find characteristic determinants for the use of ISG among this group of patients; however, the high use of an ISG in this study may confirm that an ISG is relevant for patients undergoing LSF. Socially constructed patterns of behavior may make women more prone to be contributors, and the perceived risk of spinal surgery experienced by participants with symptoms of anxiety may likewise make them more prone to ask questions and contribute actively. A total of 7 thematic categories emerged when exploring the content of posts and comments on the ISG: social recognition, experience of pain or use of pain medication, experience of physical activity or physical rehabilitation, expression of psychosocial well-being, and advising on and exploring the ISG and employment. The nature of posts and the thematic categories seem to correspond well to the prevalent occurrence of symptoms of anxiety and depression within the group of participants. Future studies need to focus on the perceived value of the ISG in LSF patients and on the usage behavior in an ISG when incorporating other features, such as animated information or instructions on a joint website.

## Acknowledgments

The authors are grateful to VISIKON for a very rewarding collaboration and highly valuable contributions in developing the Web-based platform. Also, sincere thanks to all the clinicians, patients, and their relatives who contributed in the development of the platform. Finally, a kind and sincere thanks to the facilitators, who made the platform seem alive when the first participants arrived.

## Conflicts of Interest

None declared

## References

1. Fox S, Duggan M. Pew Internet. 2013. Health Online 2013 URL: https://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-05-15] [WebCite Cache ID 78NrkTURJ]
2. Fox S. Pew Research Center. Washington DC; 2014. The Social Life of Health Information URL: https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2019-05-15] [WebCite Cache ID 78Ns7lzxy]
3. Mo PKH, Coulson NS. Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS. Psychol Health 2012;27(4):445-459. [doi: 10.1080/08870446.2011.592981] [Medline: 21854088]
4. Mo PKH, Coulson NS. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. Cyberpsychol Behav 2008 Jun;11(3):371-374. [doi: 10.1089/cpb.2007.0118] [Medline: 18537512]
5. Coulson NS. How do online patient support communities affect the experience of inflammatory bowel disease? An online survey. JRSM Short Rep 2013;4(8):2042533313478004 [FREE Full text] [doi: 10.1177/2042533313478004] [Medline: 24040493]
6. Huber J, Muck T, Maatz P, Keck B, Enders P, Maatouk I, et al. Face-to-face vs. online peer support groups for prostate cancer: a cross-sectional comparison study. J Cancer Surviv 2017 Aug 31:1-9. [doi: 10.1007/s11764-017-0633-0] [Medline: 28861845]
7. Falisi AL, Wiseman KP, Gaysynsky A, Scheideler JK, Ramin DA, Chou WS. Social media for breast cancer survivors: a literature review. J Cancer Surviv 2017 Dec;11(6):808-821. [doi: 10.1007/s11764-017-0620-5] [Medline: 28601981]
8. Algtewi E, Owens J, Baker SR. Online support groups for head and neck cancer and health-related quality of life. Qual Life Res 2017 Sep;26(9):2351-2362 [FREE Full text] [doi: 10.1007/s11136-017-1575-8] [Medline: 28417218]
9. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group for women with primary breast cancer. Cancer 2003 Mar 1;97(5):1164-1173 [FREE Full text] [doi: 10.1002/cncr.11174] [Medline: 12599221]
10. Fogel J, Albert SM, Schnabel F, Ditkoff BA, Neugut AI. Internet use and social support in women with breast cancer. Health Psychol 2002 Jul;21(4):398-404. [Medline: 12090683]
11. Kalichman SC, Cain D, Cherry C, Pope H, Eaton L, Kalichman MO. Internet use among people living with HIV/AIDS: coping and health-related correlates. AIDS Patient Care STDS 2005 Jul;19(7):439-448. [doi: 10.1089/apc.2005.19.439] [Medline: 16053401]
12. Høybye MT, Dalton SO, Deltour I, Bidstrup PE, Frederiksen K, Johansen C. Effect of internet peer-support groups on psychosocial adjustment to cancer: a randomised study. Br J Cancer 2010 Apr 27;102(9):1348-1354 [FREE Full text] [doi: 10.1038/sj.bjc.6605646] [Medline: 20424614]
13. Owen JE, Klapow JC, Roth DL, Shuster JL, Bellis J, Meredith R, et al. Randomized pilot of a self-guided internet coping group for women with early-stage breast cancer. Ann Behav Med 2005 Aug;30(1):54-64. [doi: 10.1207/s15324796abm3001_7] [Medline: 16097906]
14. Strøm J, Nielsen C, Jørgensen LB, Andersen NT, Laursen M. A web-based platform to accommodate symptoms of anxiety and depression by featuring social interaction and animated information in patients undergoing lumbar spine fusion: a

randomized clinical trial. Spine J 2019 May;19(5):827-839 [FREE Full text] [doi: 10.1016/j.spinee.2018.11.011] [Medline: 30500464]

15. Danmarks Statistik. 2016. It-anvendelse i befolkningen 2016 URL: https://www.dst.dk/da/Statistik/Publikationer/VisPub?cid=29448 [accessed 2019-05-15] [WebCite Cache ID 78Nu1lUbm]
16. Internet Live Stats. 2016. United States Internet Users URL: https://www.internetlivestats.com/internet-users/us/ [accessed 2019-05-16] [WebCite Cache ID 78PIt0NoU]
17. Edwards K, Duffy RM, Kelly BD. The silver surfer: trends of internet usage in the over 65 and the potential health benefits. Ir Med J 2015 Jun;108(6):171-174. [Medline: 26182799]
18. Bujnowska-Fedak MM. Trends in the use of the Internet for health purposes in Poland. BMC Public Health 2015;15:194 [FREE Full text] [doi: 10.1186/s12889-015-1473-3] [Medline: 25886280]
19. The Next Web. The Next Web. 2017. Digital trends 2017pages of internet, mobile and social media stats URL: https://thenextweb.com/insights/2017/01/24/digital-trends-2017-report-internet/ [accessed 2019-05-15] [WebCite Cache ID 78NumWHNM]
20. Salem DA, Bogat GA, Reid C. Mutual help goes on-line. J Community Psychol 1997;25(2):189-207. [doi: 10.1002/(SICI)1520-6629(199703)25:2<189::AID-JCOP7>3.0.CO;2-T]
21. Chee A. BinaryHealth. 2007 Jun 28. Use of Internet in Healthcare URL: https://binaryhealthcare.files.wordpress.com/2007/10/use-of-internet-in-healthcare.pdf [accessed 2019-05-20] [WebCite Cache ID 78WbHNByF]
22. Rajajee SS, Bae HW, Kanim LE, Delamarter RB. Spinal fusion in the United States: analysis of trends from 1998 to 2008. Spine 2012 Jan 1;37(1):67-76. [doi: 10.1097/BRS.0b013e31820cccfb] [Medline: 21311399]
23. Weiss AJ, Elixhauser A, Andrews RM. Characteristics of Operating Room Procedures in U.S. Hospitals, 2011. Maryland, United States: Healthcare Cost and Utilization Project; 2014 Feb. URL: https://hcup-us.ahrq.gov/reports/statbriefs/sb170-Operating-Room-Procedures-United-States-2011.pdf
24. Dean J, Potts HW, Barker C. Direction to an internet support group compared with online expressive writing for people with depression and anxiety: a randomized trial. JMIR Ment Health 2016 May 17;3(2):e12 [FREE Full text] [doi: 10.2196/mental.5133] [Medline: 27189142]
25. Arts MP, Kols NI, Onderwater SM, Peul WC. Clinical outcome of instrumented fusion for the treatment of failed back surgery syndrome: a case series of 100 patients. Acta Neurochir (Wien) 2012 Jul;154(7):1213-1217 [FREE Full text] [doi: 10.1007/s00701-012-1380-7] [Medline: 22588339]
26. Christensen FB, Laurberg I, Bünger CE. Importance of the back-café concept to rehabilitation after lumbar spinal fusion: a randomized clinical study with a 2-year follow-up. Spine (Phila Pa 1976) 2003 Dec 1;28(23):2561-2569. [doi: 10.1097/01.BRS.0000097890.96524.A1] [Medline: 14652472]
27. Falavigna A, Righesso O, Teles AR, Baseggio N, Velho MC, Ruschel LG, et al. Depression Subscale of the Hospital Anxiety and Depression Scale applied preoperatively in spinal surgery. Arq Neuropsiquiatr 2012 May;70(5):352-356 [FREE Full text] [Medline: 22618787]
28. Waxman SE, Tripp DA, Flamenbaum R. The mediating role of depression and negative partner responses in chronic low back pain and relationship satisfaction. J Pain 2008 May;9(5):434-442. [doi: 10.1016/j.jpain.2007.12.007] [Medline: 18313363]
29. Oestergaard LG, Christensen FB, Nielsen CV, Bünger CE, Fruensgaard S, Sogaard R. Early versus late initiation of rehabilitation after lumbar spinal fusion: economic evaluation alongside a randomized controlled trial. Spine (Phila Pa 1976) 2013 Nov 1;38(23):1979-1985. [doi: 10.1097/BRS.0b013e3182a7902c] [Medline: 23928716]
30. Saban KL, Penckofer SM. Patient expectations of quality of life following lumbar spinal surgery. J Neurosci Nurs 2007 Jun;39(3):180-189. [Medline: 17591414]
31. Djurasovic M, Glassman SD, Dimar JR, Crawford CH, Bratcher KR, Carreon LY. Changes in the Oswestry Disability Index that predict improvement after lumbar fusion. J Neurosurg Spine 2012 Nov;17(5):486-490. [doi: 10.3171/2012.8.SPINE12614] [Medline: 22998402]
32. Statens Institut for Folkesundhed. [National Health Board]. Copenhagen: National Health board; 2014. Danskernes Sundhed - Dan Nationale Sundhedsprofil 2013 URL: https://www.sst.dk/da/udgivelser/2014/~/media/1529A4BCF9C64905BAC650B6C45B72A5.ashx [accessed 2019-05-15] [WebCite Cache ID 78NxdftCE]
33. Flachs EM, Eriksen L, Koch MB, Ryd JT, Dibba E, Skov-Ettrup L, et al. [National Health Board]. Sundhedsstyrelsen, Syddansk Universitet: Rosendahls - Schultz Grafisk; 2015. Sygdomsbyrden I Danmark - sygdomme URL: https://www.sst.dk/da/sygdom-og-behandling/~/media/00C6825B11BD46F9B064536C6E7DFBA0.ashx [accessed 2019-05-15] [WebCite Cache ID 78Nxyod2A]
34. Hargittai E, Hinnant A. Digital inequality: differences in young adults' use of the internet. Commun Res 2008 Aug 4;35(5):602-621. [doi: 10.1177/0093650208321782]
35. Høybye MT, Dalton SO, Christensen J, Ross L, Kuhn KG, Johansen C. Social and psychological determinants of participation in internet-based cancer support groups. Support Care Cancer 2010 May;18(5):553-560. [doi: 10.1007/s00520-009-0683-6] [Medline: 19579037]
36. Falchs EM, Eriksen L, Koch MB, Ryd JT, Dibba E, Skov-Ettrup L. Google. Rosendahls-Schultz Grafisk: Sundhedsstyrelsen; 2015 Oct. Sygdomsbyrden i Danmark - Oktober URL: https://tinyurl.com/y28h3pvj [accessed 2018-01-04]

37. Smedley RM, Coulson NS, Gavin J, Rodham K, Watts L. Engagement in a newly launched online support community for complex regional pain syndrome: membership growth, header analysis and introductory messages. Int J Web Based Commun 2016;12(2):147 [FREE Full text] [doi: 10.1504/IJWBC.2016.077259]
38. Jørgensen LB, Mikkelsen LR, Noe BB, Vesterby M, Uhd M, Fridlund B. The psychosocial effect of web-based information in fast-track surgery. Health Informatics J 2017 Dec;23(4):304-318. [Medline: 27229729]
39. Rogers E. Diffusion of Innovations. New York, NY: Simon & Schuster Ltd; 2003.
40. World Health Organisation. World Medical Association. 2013. Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2019-05-15] [WebCite Cache ID 78O5rAaNi]
41. UNESCO UIS. Montreal, Quebec, Canada: United Nations Educational, Scientific and Cultural Orgabization & UNESCO Institute for Statistics; 2012. International Standard Classification of Education ISCED 2011 URL: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf [accessed 2019-05-16] [WebCite Cache ID 78PIGG2Tg]
42. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [Medline: 11832252]
43. Schreier M. What is qualitative content analysis? In: Qualitative Content Analysis in Practice First Edition. Los Angeles: Sage Publications; 2012:1-8.
44. Mason J. Qualitative Researching Second Edition. London: Sage Publications; 2002.
45. Owen JE, Klapow JC, Roth DL, Tucker DC. Use of the internet for information and support: disclosure among persons with breast and prostate cancer. J Behav Med 2004 Oct;27(5):491-505. [Medline: 15675637]
46. Manierre MJ. Gaps in knowledge: tracking and explaining gender differences in health information seeking. Soc Sci Med 2015 Mar;128:151-158. [doi: 10.1016/j.socscimed.2015.01.028] [Medline: 25618604]
47. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
48. Johnson JD, Case DO. Health Information Seeking First Edition. New York, NY: Peter Lang Publishing Inc; 2012.
49. de Vries HJ, Kloek CJJ, de Bakker DH, Dekker J, Bossen D, Veenhof C. Determinants of Adherence to the Online Component of a Blended Intervention for Patients with Hip and/or Knee Osteoarthritis: A Mixed Methods Study Embedded in the e-Exercise Trial. Telemed J E Health 2017 Dec;23(12):1002-1010. [doi: 10.1089/tmj.2016.0264] [Medline: 28525310]
50. Koball AM, Jester DJ, Domoff SE, Kallies KJ, Grothe KB, Kothari SN. Examination of bariatric surgery Facebook support groups: a content analysis. Surg Obes Relat Dis 2017 Aug;13(8):1369-1375. [doi: 10.1016/j.soard.2017.04.025] [Medline: 28600115]
51. Deetjen U, Powell JA. Informational and emotional elements in online support groups: a Bayesian approach to large-scale content analysis. J Am Med Inform Assoc 2016 May;23(3):508-513. [doi: 10.1093/jamia/ocv190] [Medline: 26911815]
52. Strøm J, Bjerrum MB, Nielsen CV, Thisted CN, Nielsen TL, Laursen M, et al. Anxiety and depression in spine surgery-a systematic integrative review. Spine J 2018 Dec;18(7):1272-1285 [FREE Full text] [doi: 10.1016/j.spinee.2018.03.017] [Medline: 29649613]
53. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
54. Carron-Arthur B, Reynolds J, Bennett K, Bennett A, Griffiths KM. What's all the talk about? Topic modelling in a mental health internet support group. BMC Psychiatry 2016 Oct 28;16(1):367 [FREE Full text] [doi: 10.1186/s12888-016-1073-5] [Medline: 27793131]
55. Nimrod G. The membership life cycle in online support groups. Int J Commun 2012;6:1245-1261 [FREE Full text]
56. 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]
57. Moore JE. Chronic low back pain and psychosocial issues. Phys Med Rehabil Clin N Am 2010 Nov;21(4):801-815. [doi: 10.1016/j.pmr.2010.06.005] [Medline: 20977962]


## Abbreviations

**eHealth:** electronic health
**HADS:** Hospital Anxiety and Depression Scale
**ISG:** Internet Support Group
**LSF:** lumbar spine fusion
**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 09.01.18; peer-reviewed by J Amann, D Bossen, J Nicholls, N Coulson; comments to author 25.05.18; revised version received 20.03.19; accepted 13.04.19; published 04.07.19.*

*Please cite as:*
*Strøm J, Høybye MT, Laursen M, Jørgensen LB, Nielsen CV*
*Lumbar Spine Fusion Patients' Use of an Internet Support Group: Mixed Methods Study*
*J Med Internet Res 2019;21(7):e9805*
*URL: https://www.jmir.org/2019/7/e9805/*
*doi:10.2196/jmir.9805*
*PMID:31274113*

Original Paper

# A Novel Intelligent Scan Assistant System for Early Pregnancy Diagnosis by Ultrasound: Clinical Decision Support System Evaluation Study

Ferdinand Dhombres[1,2], MD, PhD; Paul Maurice[1,2], MSc, MD; Lucie Guilbaud[1], MSc, MD; Loriane Franchinard[1], MSc, MD; Barbara Dias[1], MSc; Jean Charlet[2,3], PhD; Eléonore Blondiaux[4], MD, PhD; Babak Khoshnood[5], MD, PhD; Davor Jurkovic[6], MD; Eric Jauniaux[6], MD, PhD; Jean-Marie Jouannic[1,2], MD, PhD

[1]Service de Médecine Fœtale, Sorbonne Université, Assistance Publique - Hôpitaux de Paris / Hôpitaux Universitaires Est Parisiens, Hôpital Armand Trousseau, Paris, France

[2]Medical Informatics and Knowledge Engineering for eHealth Lab, INSERM, Paris, France

[3]Direction de la Recherche et de l'Innovation, Assistance Publique - Hôpitaux de Paris, Paris, France

[4]Service de Radiologie, Sorbonne Université, Assistance Publique - Hôpitaux de Paris / Hôpitaux Universitaires Est Parisiens, Hôpital Armand Trousseau, Paris, France

[5]Obstetrical, Perinatal and Pediatric Epidemiology Research Team, Center for Biostatistics and Epidemiology, INSERM, Paris, France

[6]Gynaecology Diagnostic and Outpatient Treatment Unit, University College Hospital and Institute for Women's Health, University College London, London, United Kingdom

**Corresponding Author:**
Ferdinand Dhombres, MD, PhD
Service de Médecine Fœtale
Sorbonne Université
Assistance Publique - Hôpitaux de Paris / Hôpitaux Universitaires Est Parisiens, Hôpital Armand Trousseau
26 avenue du Dr Arnold Netter
Paris, 75012
France
Phone: 33 622286740
Email: ferdinand.dhombres@inserm.fr

## Abstract

**Background:** Early pregnancy ultrasound scans are usually performed by nonexpert examiners in obstetrics/gynecology (OB/GYN) emergency departments. Establishing the precise diagnosis of pregnancy location is key for appropriate management of early pregnancies, and experts are usually able to locate a pregnancy in the first scan. A decision-support system based on a semantic, expert-validated knowledge base may improve the diagnostic performance of nonexpert examiners for early pregnancy transvaginal ultrasound.

**Objective:** This study aims to evaluate a novel Intelligent Scan Assistant System for early pregnancy ultrasound to diagnose the pregnancy location and determine the image quality.

**Methods:** Two trainees performed virtual transvaginal ultrasound examinations of early pregnancy cases with and without the system. The ultrasound images and reports were blindly reviewed by two experts using scoring methods. A diagnosis of pregnancy location and ultrasound image quality were compared between scans performed with and without the system.

**Results:** Each trainee performed a virtual vaginal examination for all 32 cases with and without use of the system. The analysis of the 128 resulting scans showed higher quality of the images (quality score: +23%; $P$<.001), less images per scan (4.6 vs 6.3 [without the CDSS]; $P$<.001), and higher confidence in reporting conclusions (trust score: +20%; $P$<.001) with use of the system. Further, use of the system cost an additional 8 minutes per scan. We observed a correct diagnosis of pregnancy location in 39 (61%) and 52 (81%) of 64 scans in the nonassisted mode and assisted mode, respectively. Additionally, an exact diagnosis (with precise ectopic location) was made in 30 (47%) and 49 (73%) of the 64 scans without and with use of the system, respectively. These differences in diagnostic performance (+20% for correct location diagnosis and +30% for exact diagnosis) were both statistically significant ($P$=.002 and $P$<.001, respectively).

**Conclusions:** The Intelligent Scan Assistant System is based on an expert-validated knowledge base and demonstrates significant improvement in early pregnancy scanning, both in diagnostic performance (pregnancy location and precise diagnosis) and scan quality (selection of images, confidence, and image quality).

***(J Med Internet Res 2019;21(7):e14286)*** doi:10.2196/14286

**KEYWORDS**

decision support system; ontology; knowledge base; medical ultrasound; ectopic pregnancy

## Introduction

### Background

Ectopic pregnancy (EP) is defined by implantation of the gestational sac outside the endometrial cavity and occurs in 1%-2% of all pregnancies [1]. EP accounts for approximately 3%-5% of pregnancy-related deaths in developed countries [2]. Around 95% of EPs implant in the fallopian tubes and 5%-7% implant within the uterine wall but outside the uterine cavity. Nontubal EPs are more difficult to diagnose than tubal EP and are associated with a higher mortality and morbidity [3]. Delayed diagnosis is the main factor for EP associated with maternal death [4] and affects the success rate of future pregnancies [5]. Skilled ultrasound operators can diagnose an EP at an early stage by transvaginal sonography (TVS), often at the first examination [6]. However, less experienced operators perform first-line screening for patients at risk of EP in most emergency units; for them, this diagnosis remains difficult and more than three examinations are often needed [7].

### Prior Work

We have developed the first Intelligent Scan Assistant System for early pregnancy TVS examination. This clinical decision support system (CDSS) [8,9] is a computer program that provides diagnosis assistance during TVS examination of pregnancy of unknown location. During an ultrasound examination and in real time, this system assists the operator by suggesting ultrasound views to acquire and relevant signs to look for; it also displays reference ultrasound images demonstrating these relevant signs and views (from expert-reviewed collections of early pregnancy cases). The semantic design and features of this CDSS have been published in the medical semantics informatics community [10]. One key feature of this system is the personalized imaging protocol [11]: The system guides the operator through a structured acquisition of decisive ultrasound images to optimize the diagnostic pathway. We deemed this system “intelligent” because these personalized imaging protocols are not precalculated, but dynamically derived by the system (by SPARQL queries on the early pregnancy ontology of the knowledge base) from the guided image analysis of the current case.

In a preliminary study, this novel system demonstrated efficient support for a precise ultrasound image analysis, with a precision of 83% for the identification of signs in a series of 208 retrospectively collected ultrasound images of various types of ectopic pregnancies [10,11].

### Study Goal

In this study, we aimed to assess the added value of this novel CDSS for early pregnancy ultrasound. Our objective was to evaluate the effect of using the CDSS during TVS on scan quality and accuracy of the diagnosis of pregnancy location.

## Methods

### Clinical Decision Support System Evaluation Overview

Two obstetrics and gynecology (OB/GYN) trainees with basic national training in ultrasound imaging (including early pregnancy courses and simulation sessions) performed 32 ultrasound examinations in early pregnancy patients without and with the CDSS. These were re-examinations of prospectively collected 3D volumes from ultrasound data from the gynecology emergency unit at a university hospital. At the beginning of the study, the two trainees viewed a 2-minute video presentation of the CDSS and had a 10-minute hands-on session with the team who developed the CDSS. The TVS of early pregnancy cases was performed using a simulation device with and without the CDSS.

### Ultrasound (Transvaginal) Simulator and 3D Ultrasound Volume Collection

The simulation device was the interpolative model-based ultrasound simulator ScanTrainer (MedaPhor, Wales, United Kingdom) with a realistic haptic feedback transvaginal probe. This ultrasound simulator produces 2D images generated from 3D vaginal ultrasound volumes, which had been acquired during actual vaginal scans [12]. The complete virtual TVS platform used for the study integrates the ultrasound simulator and the CDSS with a dual screen setting (Figure 1). One screen displays the usual information for scanning as a regular ultrasound system. The other screen displays the CDSS interface for the image analysis and scan assistance (Figure 2).

Thirty-two 3D vaginal ultrasound volumes for this study were collected from patients during early pregnancy emergency examinations in one university hospital center (Figure 3), using an expert 3D ultrasound system (GE Healthcare Voluson E10/E8 with a RIC5-9-D vaginal probe, Cincinnati, OH). In our center, the first ultrasound examinations are always performed by junior OB/GYN examiners. In case of pregnancy of unknown locations or EP, a second vaginal scan is performed by a senior OB/GYN examiner. We collected 16 consecutive cases of pregnancy of unknown locations and 16 consecutive cases of EP diagnosed after the first ultrasound examination. For this study, three additional 3D volume acquisitions were performed by the senior OB/GYN examiner during the second TVS (acquisition field of 180°/100°): one volume for the uterus and one adjacent volume for each adnexal region. In case of a suspected ruptured EP, the volume acquisition was not performed, to avoid any delay in performing the surgical procedure. Additionally, when the diagnosis of intrauterine pregnancy was obvious (normal pregnancy of 6 weeks of gestation or more) at the second examination by the senior, no 3D volume was acquired. Rare types of ectopic pregnancy (heterotopic pregnancy, interstitial pregnancy, caesarean-section scar pregnancy, and cervical pregnancy) were also excluded from this study. Thus, a consecutive series of 32 sets of 3D vaginal ultrasound volumes was collected, deidentified, and imported in the ultrasound simulator. The final diagnoses of the 32 cases in this series were intrauterine pregnancy (n=18) and tubal EP (n=14), all correctly diagnosed by senior TVS experts and confirmed by pregnancy outcomes.

**Figure 1.** Global view of the virtual vaginal ultrasound platform for evaluation of the Intelligent Scan Assistant System. The left monitor displays the ultrasound simulator interface (ScanTrainer, MedaPhor, Wales, United Kingdom) and the right monitor displays the decision support system (Intelligent Scan Assistant System).

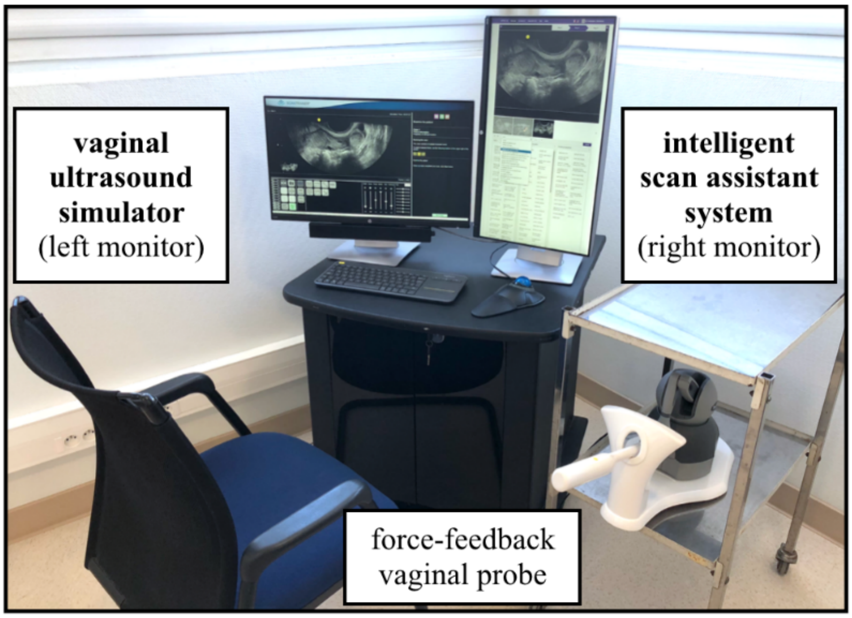



**Figure 2.** Detailed view of the Intelligent Scan Assistant System (right monitor). The two main steps with the decision support system on the right monitor are image analysis and scan assistance.

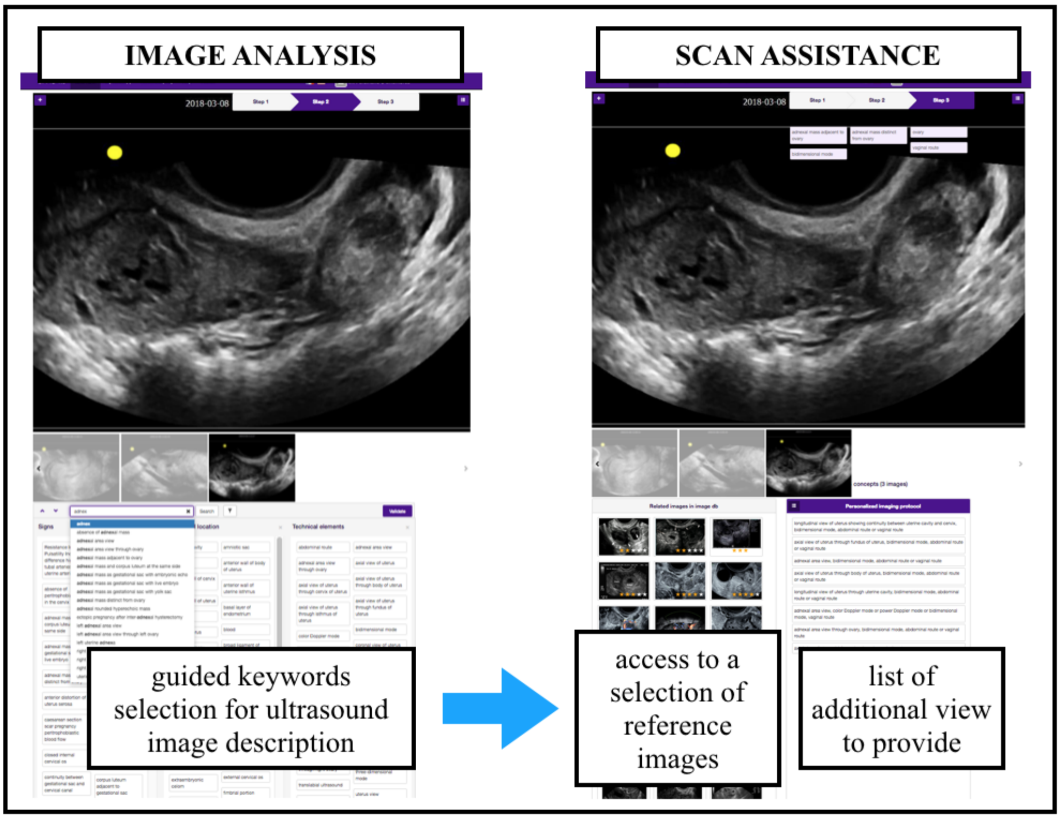
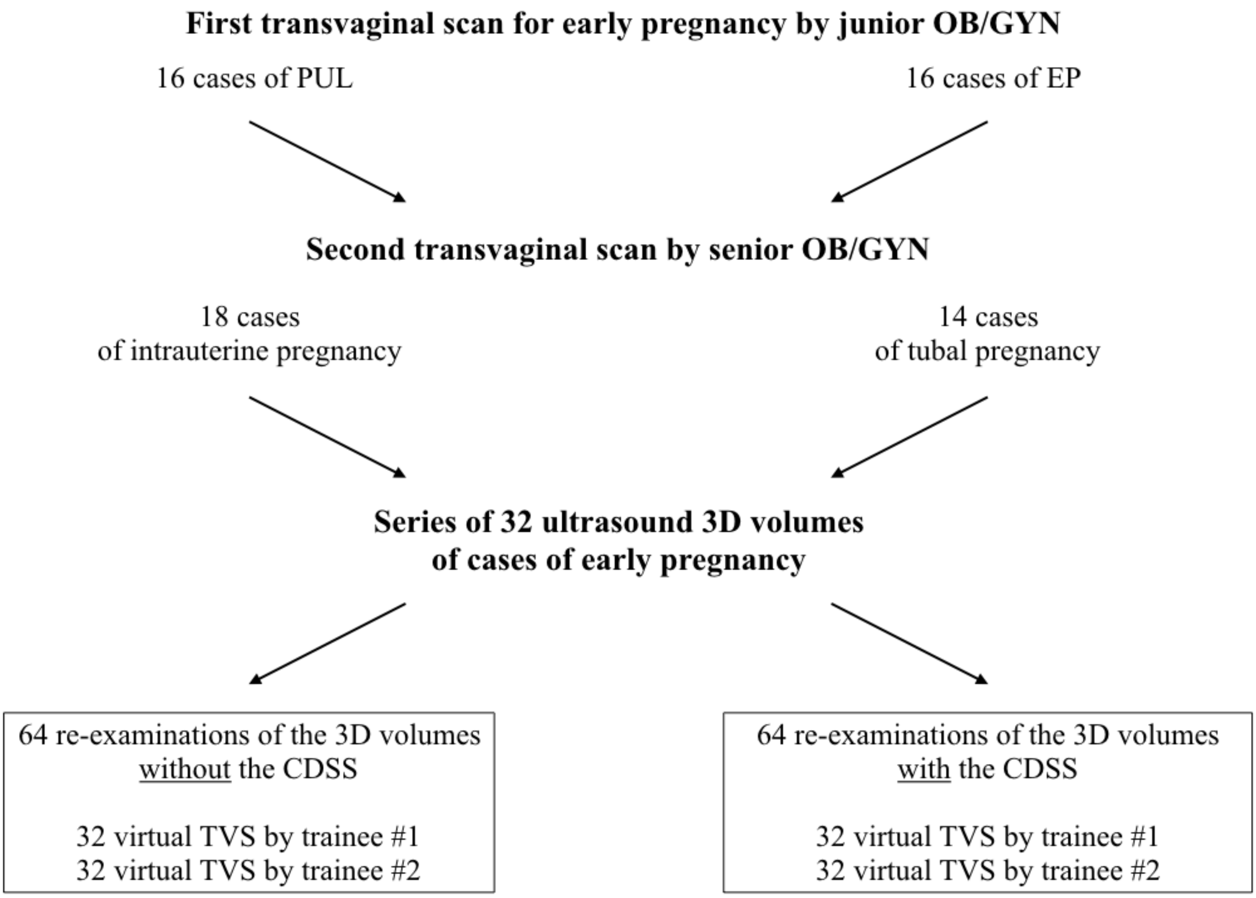


**Figure 3.** Three-dimensional ultrasound volume acquisition flowchart. Selection of cases for the 3D ultrasound volume series used for the virtual transvaginal scans (TVS) in this study. CDSS: clinical decision support system; EP: ectopic pregnancy; OB/GYN: obstetrics and gynecology; PUL: pregnancy of unknown locations.

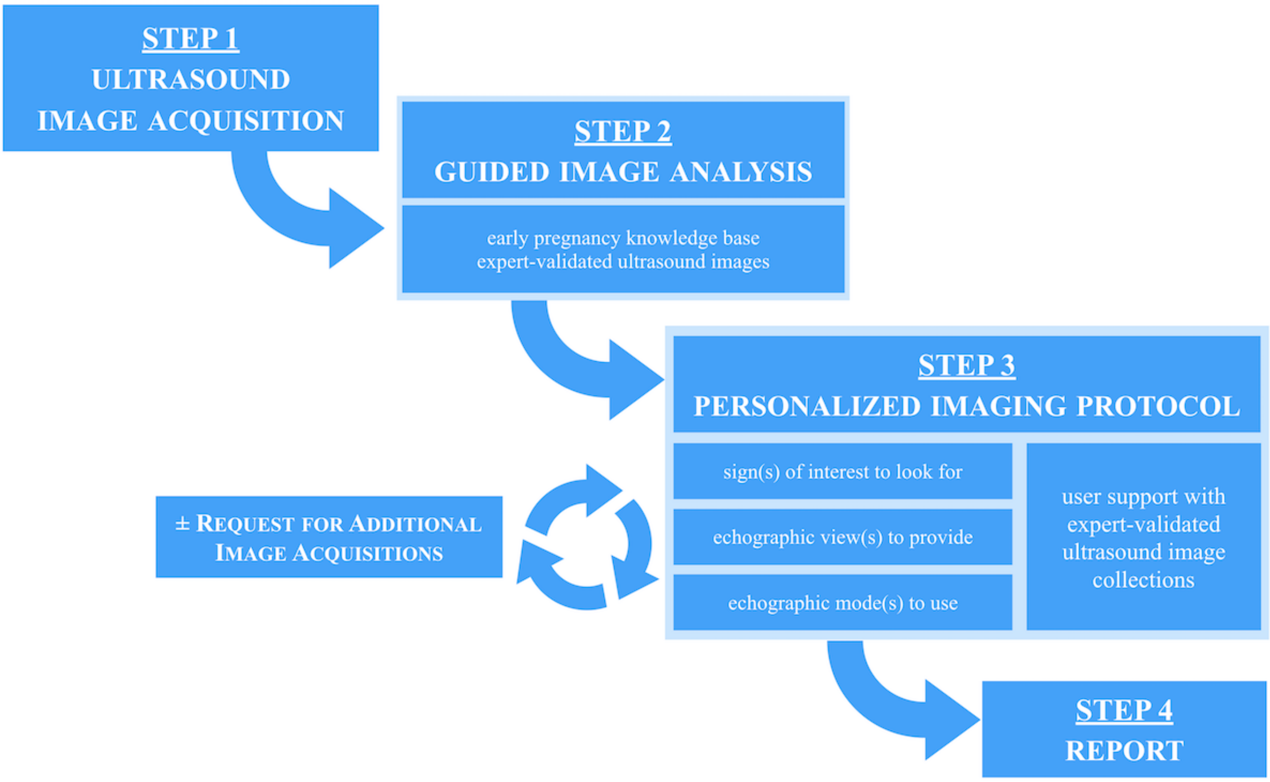



## Clinical Decision Support System Evaluation Protocol

### *Virtual Ultrasound Examinations*

The two trainees performing the virtual TVS were independent of acquisition of 3D volumes, and they were unaware of the medical report and final diagnosis. The clinical information provided for the scans were identical for all cases and limited to “moderate pelvic pain and positive pregnancy test.” The 32 scans were performed twice by each trainee without supervision in a random order in a nonassisted mode (without the CDSS) and 2 months later in assisted mode (with the CDSS). The potential recall bias was also controlled by the 2-month interval between the two TVS sessions. Additionally, it should be mentioned that during these 2 months, the two trainee operators did not receive any ultrasound training and did not have any ultrasound scanning activity. In the nonassisted mode, the scans were performed following the usual protocol for OB/GYN emergency ultrasound in our center [13,14], using a standardized reporting system. In the assisted mode, the scans were performed following a personalized image analysis and acquisition protocol suggested by the CDSS [10,11]. The personalized imaging protocol and the CDSS workflow are presented in Figure 4. Briefly, in step 1, the operator performs the scan and acquires ultrasound images. In step 2, he/she follows the system guidance for a precise analysis of these acquired images: He/she describes the image with keywords for anatomical structures, ultrasound signs, and technical elements (ultrasound route, mode, and view). The keywords are displayed with text definitions and are illustrated by expert-validated images. In step 3, if necessary, the system suggests providing additional imaging elements (ultrasound views, signs, and anatomical structures), thus assisting the operator in establishing a comprehensive image set in order to reach a precise diagnosis. This is a personal imaging protocol that is automatically calculated by the system. The personal imaging protocol is derived from computer-based reasoning over the early pregnancy knowledge base. After step 3, the user may either follow the personal imaging protocol and provide the additional requested elements or proceed to the final report and finish the examination (step 4).

**Figure 4.** Personalized imaging protocol and workflow of the Intelligent Scan Assistant System for ultrasound imaging.




### *Ultrasound Images and Report Scoring Methods*

For each examination, we collected the images, reports, and data on duration of the scans. Two senior experienced ultrasound operators reviewed the images and reports of the virtual examinations. During the review, they were blinded to the use of the CDSS and the final diagnosis. They had the same minimal clinical information for all cases as the two trainees: "moderate pelvic pain and positive pregnancy test."

They scored the images according to the quality criteria for the sagittal view of the uterus and the ovaries as per a previous study [14]. The maximum quality score was 15 points (Textbox 1). They also performed a subjective scoring of the scans and reports, reflecting their trust in the conclusion of the report associated with the images. This level of trust was assessed with a 5-level scale (Textbox 2).

**Textbox 1.** The image set quality was assessed using a score based on 15 items.

**On the sagittal view of the uterus, the five quality criteria were as follows:**

- "uterine cervix visible" (1 point)
- "uterine fundus visible" (1 point)
- "endometrial midline echo visible" (1 point)
- "endocervix visible" (1 point)
- "uterus occupying more than half of the total image size" (1 point)

**On each of the views of the ovaries, the 5×2 quality criteria were as follows:**

- "side stated" (1 point)
- "follicle(s) visible" (1 point)
- "iliac vein visible" (1 point)
- "long axis of the ovary <30° with the horizontal line" (1 point)
- "vary occupying more than a quarter of the total image size" (1 point)

**Textbox 2.** The level of trust in the report was assessed using a 5-level scale.

- **Level 1: No trust in the final diagnosis (incorrect):** The diagnosis is most likely incorrect, and the image set suggests another diagnosis. Immediate supervisor examination is needed.
- **Level 2: No trust in the final diagnosis (low quality):** The image set quality is insufficient and/or does not support the final diagnosis. Immediate supervisor examination is needed.
- **Level 3: Moderate trust in the final diagnosis:** Although the diagnosis might be correct, the image set quality is insufficient, and a supervisor examination is needed.
- **Level 4: The image set quality could be improved**; however, it is of sufficient quality to accurately support the final diagnosis. No supervisor examination is needed.
- **Level 5 represents a total trustin the final diagnosis**: The image set effectively supports the diagnosis, and no supervisor examination is needed.

### Statistical Analysis

The reproducibility of the scoring methods for quality and trust was assessed on a random sample of 20% of all scans (n=25) and independently reviewed by both experts. We tested for the differences in trust and quality scores. We also tested for the differences between examination modes (assisted vs nonassisted mode) in the diagnosis of location of pregnancy (ectopic OR nonectopic) and in the final diagnosis precision (exact location of the ectopic pregnancy, ie, “tubal pregnancy” explicitly stated in the report conclusion). The gold standard for the diagnosis was the final diagnosis in all cases, as reported in the senior TVS report and confirmed by the pregnancy follow-up.

Statistical analysis was performed using R, version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria) and STATA, version 15 (StataCorp, College Station, TX). Paired *t* tests were performed to compute the difference in continuous variables (scan duration, image count, quality score, and trust score). Exact McNemar tests were used to test for the differences in categorical variables (presence of the three mandatory ultrasound views, diagnosis of location, and final diagnosis precision). We also calculated differences in the proportions of outcomes, with 95% CIs, for assisted versus nonassisted modes. Adjusted kappa coefficients (Cohen weighted kappa) for quality scores and trust scores were computed to test for the reproducibility of the scoring methods.

For all tests, a $P$ value $\leq$.05 was considered statistically significant. Adjusted kappa values <0.6, between 0.6 and 0.8, and >0.8 were considered to represent poor, moderate, and good agreement, respectively.

### Ethics Approval

The development of this CDSS for early pregnancy (including expert-validated early pregnancy ultrasound images) and the evaluation study (including collection and analysis of 3D ultrasound volumes of early pregnancy) were both approved by the French National College of the OB/GYN Institutional Review Board (CNGOF Research Ethics Committee CEROG #2015-GYN-1002 and #2016-GYN-0601, respectively).

## Results

### Virtual Scans and Scoring Method Reproducibility

Each trainee performed a virtual transvaginal examination for all 32 cases with and without the system. The expert operators reviewed the 128 resulting scans for quality of images and trust in the reports. The experts’ agreement was tested on 25 scans. The level of agreement was good, with kappa values of 0.86 (0.76-0.96) for objective quality scoring and 0.86 (0.70-1.0) for subjective trust scoring.

### Impact of the Clinical Decision Support System on Image Quality

The scan quality differences are presented in Table 1. We found that the average quality score for ultrasound images was 23% higher when using the CDSS than with the nonassisted mode, with an average value of 12.6 of 15 ($P$<.001). Additionally, when using the CDSS, the number of images per scan was lower than that with the nonassisted mode (4.6 vs 6.3, $P$<.001).

The average trust score was 20% higher when using the CDSS than with the nonassisted mode, with an average value of 4.12 of 5 ($P$<.001).

**Table 1.** Differences in scan quality with (assisted mode) and without (nonassisted mode) the decision support system.[a]

| Scan quality parameter | Assisted mode (64 scans) | Nonassisted mode (64 scans) | Difference | *P* value |
|---|---|---|---|---|
| Image count in report, mean (SD) | 4.64 (0.80) | 6.33 (2.07) | –1.69 (–27%) | <.001 |
| Scan duration (minutes), mean (SD) | 14.7 (7.1) | 6.4 (3.3) | +8.3 (+129%) | <.001 |
| Quality score of image sets, mean (SD) | 12.5 (1.86) | 10.2 (1.90) | +2.3 (+23%) | <.001 |
| Trust score of report, mean (SD) | 4.12 (0.83) | 3.42 (1.04) | +0.70 (+20%) | <.001 |

[a]The tests for difference were paired *t* tests.

### Impact of the Clinical Decision Support System on Diagnosis of Pregnancy Location

The diagnosis differences are displayed in Table 2. We observed a correct diagnosis of location in 39 (61%) and 52 (81%) of 64 scans in the nonassisted mode and assisted mode, respectively. Additionally, the exact diagnosis was achieved in 30 (47%) and 49 (77%) scans in the nonassisted mode and assisted mode, respectively. These differences (+20% for correct location diagnosis and +30% for exact diagnosis) were both statistically significant ($P$=.002 and $P$<.001, respectively).

Without the use of the CDSS, we recorded 8 false-negative diagnoses of tubal EP (cases 44 and 23 for both trainees and cases 50, 45, 33, and 1 for one trainee). With the CDSS, the false-negative result for ectopic pregnancy was a scan of a tubal pregnancy case (case 44 for one trainee). In the other seven cases with false-negative diagnoses of EP, all relevant signs associated with the final diagnosis (with reference images) were presented to the trainee when using the CDSS. More precisely, images exhibiting the key features of tubal pregnancy were acquired following the personalized protocol of the CDSS and correctly diagnosed. Additionally, the quality score, trust score, and number of images per scan were significantly different when using the CDSS as compared to not using the CDSS: 12.6 versus 10.2 ($P$=.01), 4.0 versus 3.0 ($P$=.02), and 4.7 versus 6.5 ($P$=.04), respectively. The scan duration was also significantly different when using the CDSS as compared to not using the CDSS (14.4 versus 7.6 min; $P$=.003).

**Table 2.** Differences in the diagnostic performance of trainees with (assisted mode) and without the decision support system (nonassisted mode).[a]

| Diagnostic performance parameter | Assisted mode (64 scans), n (%) | Nonassisted mode (64 scans), n (%) | Difference, n | Difference, % (95% CI) | $P$ value |
|---|---|---|---|---|---|
| Correct pregnancy location (ectopic/nonectopic) | 52 (81) | 39 (61) | 13 | +20 (7-33) | .002 |
| Exact diagnosis (with precise ectopic location) | 49 (77) | 30 (47) | 19 | +30 (15-44) | <.001 |
| False-negative of ectopic pregnancy | 1 (1.6) | 8 (12.5) | –7 | –10.9 | —[b] |
| False-positive of ectopic pregnancy | 3 (4.7) | 3 (4.7) | 0 | — | — |

[a]The test for difference was exact McNemar test.

[b]Not available.

We observed 3 false-positive EP diagnoses in the assisted mode and 3 other false-positive EP diagnoses in the nonassisted mode, which were the 6 cases of intrauterine pregnancies.

## Discussion

### Principal Results

Our study demonstrated a significant improvement in early pregnancy ultrasound examination, both in diagnostic performance (pregnancy location and precise diagnosis) and scan quality (selection of images, confidence, and image quality) for OB/GYN trainees using the CDSS, when pregnancy of unknown locations or EP was suspected.

Definitive diagnosis of ectopic pregnancy can be achieved by TVS, but it relies on a precise analysis of ultrasound findings [1,15-19]. However, in most emergency units, the initial TVS is usually performed by trainees or sonographers with basic expertise in OB/GYN scanning. The support of expert-validated images in addition to the personalized protocol (intelligent suggestions of ultrasound signs, views, and modes) are key features of the CDSS, especially for improving the false-negative diagnoses of EP. During the examination, it provides actionable knowledge to less experienced operators, thus improving their diagnostic capabilities in real time. Our results suggest that the CDSS improves not only the diagnosis of early pregnancy location, but also the diagnostic accuracy. The use of the CDSS resulted in a better selection of images with higher quality. This facilitates the review of the scans by the senior experts in our department, as suggested by their higher trust scores.

### Limitations

The main limitation of our study is that our evaluation relies on virtual TVS. In a previous study, Infantes et al [20] showed that offline analysis of 3D TVS static datasets has limitations in terms of the diagnostic accuracy for EP [20]. However, their study was not conducted with ultrasound simulation platforms [12,21]. In our study, we chose the best simulation options for realistic 2D ultrasound examinations. This led to a moderate loss in image quality, but with this study design, the same patient would have been scanned twice (with and without the CDSS) and by each trainee (32 patients, 128 scans; 4 scans per patient), thus specifically assessing the potential added value of the CDSS itself. A key skill in ultrasound is to find the pathology and, in the simulator, the trainees were presented with volumes that contained all necessary information to make a diagnosis; therefore, their scanning skills were not properly evaluated in this study. However, even if the interpretation was easier, we believe this was not a bias in favor of the CDSS. Interestingly, the trainees complained about the lack of color Doppler imaging (CDI) in the simulator, but only when they used the CDSS and, in particular, for the cases of false-positive diagnoses of EP. Better ultrasound imaging quality and access to the CDI mode might change the performances of the CDSS. As the CDSS includes a rich CDI semiology of EP, this change might even be in favor of the CDSS. Palpation by a transvaginal probe often provides critical information to make the correct diagnosis. This cannot be done on current simulators, which is another limitation of the study.

We observed an increase of 8 minutes in the scan duration. Similar additional time costs were observed in a pilot study when using standardized protocols with integrated software for the second trimester screening [22]. Consequently, we believe that technical improvements, in particular, integration in the ultrasound platform, could improve examination durations; in our study, each image file was manually imported in the CDSS during the TVS. Overall, an additional 8 minutes is a reasonable cost for significant diagnostic improvements, and consequently, for a reduction in the number of visits to diagnose the correct location of the pregnancy.

### Comparison With Prior Work

This CDSS is the first computer-based reasoning system in the field of OB/GYN. In the current trend of new technical solutions to improve ultrasound examinations, this CDSS has been evaluated, offers novel intelligent scan assistance in real time, is dynamically based on previous ultrasound findings, and has salient reference images in backup. In contrast, only 58% of CDSSs demonstrated improvements in the processes of care [23].

Software-enforced standardized protocols for screening offer interesting solutions for ultrasound scan improvement [22]. These systems implement static checklists to improve the acquisition of standard image sets. Other tools automated the 2D images processing (eg, for caliper positioning) and the 3D/4D volume processing (eg, to derive 2D images of the fetal brain and heart) [24-26]. Finally, online resources provide access to collections of medical images, including ultrasound and OB/GYN material [27-29].

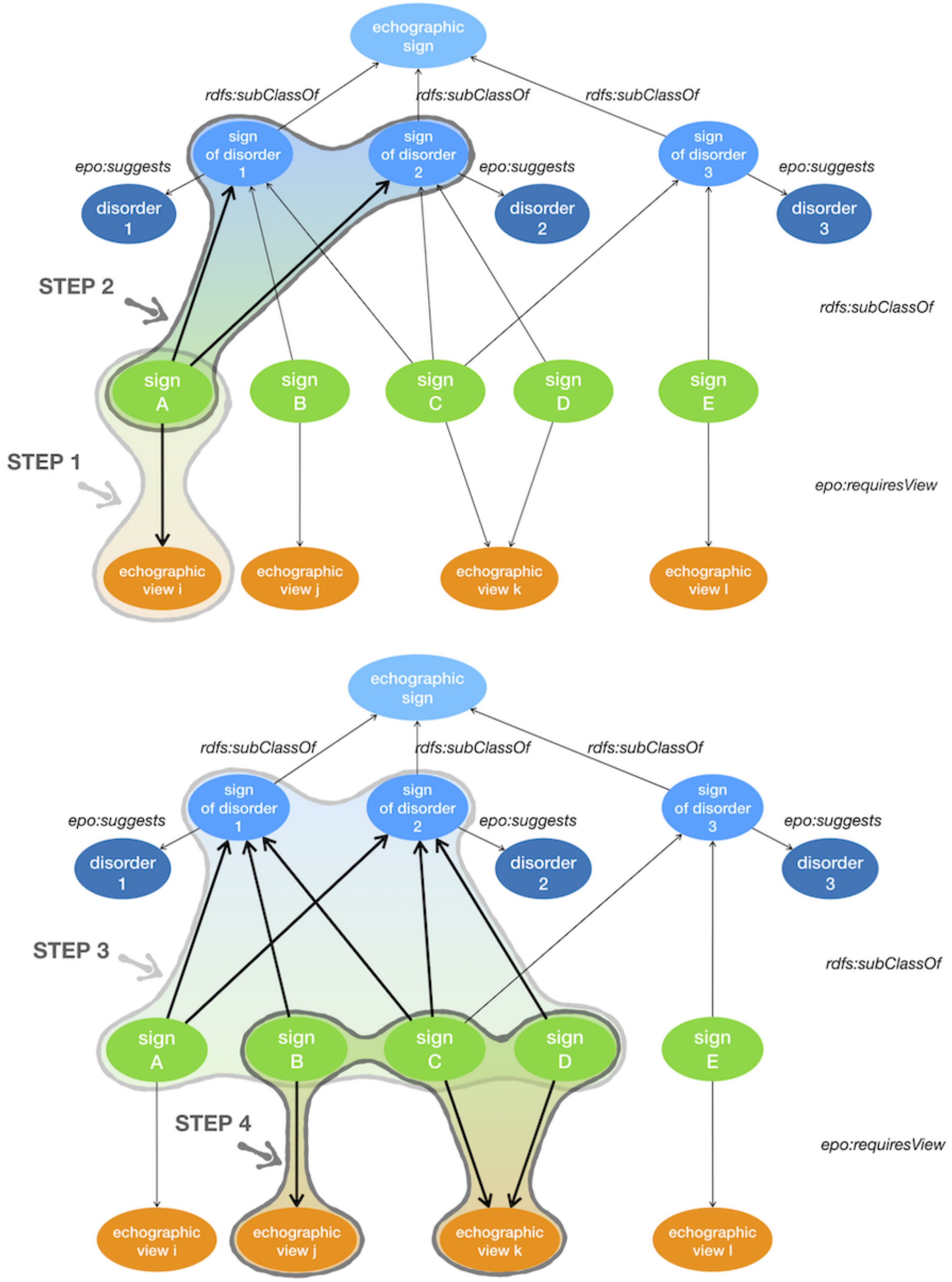
Our choice of a CDSS based on ontology and semantic Web technologies have several significant advantages. The ontology is a model representing the medical knowledge involved in TVS for early pregnancy. This model allows computer-based reasoning and enables the personalized imaging protocol feature of the CDSS. Of particular interest, this type of ontology-based reasoning CDSS differs from current systems (eg, deep learning and neural network systems) and does not integrate any “black-box” component [9]: Every step of the calculation can be audited and is readable by a human. More specifically, Figure 5 illustrates the effective support of the ontology to derive a personalized imaging protocol. Every step of the protocol relies on SPARQL queries to navigate through the graph of the knowledge base (XML/RDF). This knowledge base is present in a triplestore with semantic inference capabilities (OWL/HermIT). Consequently, the result of every step of the protocol is a set of resource description framework triple, with labels (skos:prefLabel) that can be reviewed by medical experts.

When a sign is identified during the scan, in an echographic view, the clinical reasoning principle is formalized as follows:

1. List all disorders suggested by the identified sign(s).
2. For all these disorders, list all associated sign(s).
3. For all these signs, list all required echographic view(s).
4. Provide support to the operator: ordered list of echographic views required to look for relevant signs for differential diagnosis.

The CDSS design implements international standards (RDF, SPARQL, and OWL) with a generic strategy for medical imaging. The CDSS was initially developed for early pregnancy ultrasound. There is no technical obstacle to extending the system to other areas of ultrasound imaging such as diagnosis of placentation disorders or morphological ultrasound examination of the fetus. Furthermore, it allows for a simple curation process (eg, addition of new signs or new cases in the collection) and does not require specific skills in informatics. For example, when new ultrasound imaging features are described in medical publications (eg, when fetal ultrasound features of postnatal disorders are discovered, as it was recently the case for the limited dorsal myeloschisis, which is a well-known postnatal disorder [30]), updating the whole system is easy. In contrast, updating usual expert systems would require technical developments. Finally, semantic Web technologies are designed to scale and support interoperability. The scaling capability is the capability to handle a large amount of data and is a prerequisite to cover the large domain of fetal abnormalities, including ultrasound features, anatomical locations, adequate ultrasound views, and nosology of fetal disorders. The interoperability capability opens data integration with other databases, in particular, genetic data repositories. This interoperability is a possible way to establish correlations between ultrasound phenotypes and genetic variants [31].

**Figure 5.** Principles of clinical reasoning for the CDSS represented in the ontology for early pregnancy (epo). Step 1: Identification of "epo:sign_A", using "epo:echographic_view_i". Step 2: Compute the list of disorders suggested by "epo:sign_A," "epo:disorder_1," and "epo:disorder_2". Step 3: Compute the list of signs for the list of disorders: "epo:sign_B," "epo:sign_C," and "epo:sign_D". Step 4: Suggest a list of echographic views required for the list of signs: "epo:echographic_view_j" and "epo:echographic_view_k". CDSS: clinical decision support system.




## Conclusions

In the growing ecosystem of emerging new tools for medical imaging assistance, the Intelligent Scan Assistant System is a CDSS based on a semantic representation of expert knowledge, consistent with a complementary solution that promotes improvement of both scan quality and diagnostic accuracy. The evaluation of the system on a simulation platform demonstrated its added value for trainees in TVS. Consequently, the implementation of this CDSS in routine care may reduce the number of TVS examinations to the minimum number of TVSs required to diagnose (or exclude) EP. These results await confirmation by randomized control trials and further use in different areas of OB/GYN imaging.

## Acknowledgments

We thank Dr Nicolas Perrot (Medical Imaging Center, Pyramides, Paris, France) for his help during the ultrasound volume acquisition. The SATT-Lutech and Sorbonne University, Paris, France, supported and developed of the Intelligent Scan Assistant System prototype, and the AP-HP founded the ultrasound simulation platform used in this study.

## Conflicts of Interest

None declared.

## References

1. Memtsa M, Jamil A, Sebire N, Jauniaux E, Jurkovic D. Diagnosis and management of intramural ectopic pregnancy. Ultrasound Obstet Gynecol 2013 Sep;42(3):359-362 [FREE Full text] [doi: 10.1002/uog.12437] [Medline: 23417903]
2. Creanga AA, Syverson C, Seed K, Callaghan WM. Pregnancy-Related Mortality in the United States, 2011-2013. Obstet Gynecol 2017 Dec;130(2):366-373 [FREE Full text] [doi: 10.1097/AOG.0000000000002114] [Medline: 28697109]
3. Knez J, Day A, Jurkovic D. Ultrasound imaging in the management of bleeding and pain in early pregnancy. Best Pract Res Clin Obstet Gynaecol 2014 Jul;28(5):621-636. [doi: 10.1016/j.bpobgyn.2014.04.003] [Medline: 24841987]
4. Cantwell R, Clutton-Brock T, Cooper G, Dawson A, Drife J, Garrod D, et al. Saving Mothers' Lives: Reviewing maternal deaths to make motherhood safer: 2006-2008. The Eighth Report of the Confidential Enquiries into Maternal Deaths in the United Kingdom. BJOG 2011 Mar;118 Suppl 1:1-203 [FREE Full text] [doi: 10.1111/j.1471-0528.2010.02847.x] [Medline: 21356004]
5. Ghaneie A, Grajo JR, Derr C, Kumm TR. Unusual ectopic pregnancies: sonographic findings and implications for management. J Ultrasound Med 2015 Jun;34(6):951-962. [doi: 10.7863/ultra.34.6.951] [Medline: 26014313]
6. Jurkovic D, Wilkinson H. Diagnosis and management of ectopic pregnancy. BMJ 2011 Jun 10;342:d3397. [doi: 10.1136/bmj.d3397] [Medline: 21665933]
7. Wedderburn CJ, Warner P, Graham B, Duncan WC, Critchley HOD, Horne AW. Economic evaluation of diagnosing and excluding ectopic pregnancy. Hum Reprod 2010 Feb;25(2):328-333 [FREE Full text] [doi: 10.1093/humrep/dep397] [Medline: 19933287]
8. Shortliffe EH. Computer programs to support clinical decision making. JAMA 1987 Jul 03;258(1):61-66. [Medline: 3586293]
9. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA 2018 Dec 04;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
10. Dhombres F, Maurice P, Friszer S, Guilbaud L, Lelong N, Khoshnood B, et al. Developing a knowledge base to support the annotation of ultrasound images of ectopic pregnancy. J Biomed Semantics 2017 Jan 31;8(1):4 [FREE Full text] [doi: 10.1186/s13326-017-0117-1] [Medline: 28137311]
11. Maurice P, Dhombres F, Blondiaux E, Friszer S, Guilbaud L, Lelong N, et al. Towards ontology-based decision support systems for complex ultrasound diagnosis in obstetrics and gynecology. J Gynecol Obstet Hum Reprod 2017 May;46(5):423-429. [doi: 10.1016/j.jogoh.2017.03.004] [Medline: 28934086]
12. Chalouhi GE, Bernardi V, Ville Y. Ultrasound simulators in obstetrics and gynecology: state of the art. Ultrasound Obstet Gynecol 2015 Sep 04;46(3):255-263. [doi: 10.1002/uog.14707]
13. Popowski T, Huchon C, Fathallah K, Bouhanna P, Bernard J, Fauconnier A. Standardization of the gynecological emergency ultrasound examination. Gynecol Obstet Fertil 2012 Mar;40(3):182-188. [doi: 10.1016/j.gyobfe.2011.08.028] [Medline: 22079742]
14. Salomon LJ, Nassar M, Bernard JP, Ville Y, Fauconnier A, Société Française pour l'Amélioration des Pratiques Echographiques (SFAPE). A score-based method to improve the quality of emergency gynaecological ultrasound examination. Eur J Obstet Gynecol Reprod Biol 2009 Apr;143(2):116-120. [doi: 10.1016/j.ejogrb.2008.12.003] [Medline: 19203825]
15. Jurkovic D, Mavrelos D. Catch me if you scan: ultrasound diagnosis of ectopic pregnancy. Ultrasound Obstet Gynecol 2007 Jul;30(1):1-7 [FREE Full text] [doi: 10.1002/uog.4077] [Medline: 17587215]
16. Barnhart K, van Mello NM, Bourne T, Kirk E, Van Calster B, Bottomley C, et al. Pregnancy of unknown location: a consensus statement of nomenclature, definitions, and outcome. Fertil Steril 2011 Mar 01;95(3):857-866 [FREE Full text] [doi: 10.1016/j.fertnstert.2010.09.006] [Medline: 20947073]
17. Kirk E, Bottomley C, Bourne T. Diagnosing ectopic pregnancy and current concepts in the management of pregnancy of unknown location. Hum Reprod Update 2014;20(2):250-261. [doi: 10.1093/humupd/dmt047] [Medline: 24101604]
18. [No author listed]. ACOG Practice Bulletin No. 193: Tubal Ectopic Pregnancy. Obstet Gynecol 2018 Dec;131(3):e91-e103. [doi: 10.1097/AOG.0000000000002560] [Medline: 29470343]
19. Nadim B, Infante F, Lu C, Sathasivam N, Condous G. Morphological ultrasound types known as 'blob' and 'bagel' signs should be reclassified from suggesting probable to indicating definite tubal ectopic pregnancy. Ultrasound Obstet Gynecol 2018 Apr;51(4):543-549 [FREE Full text] [doi: 10.1002/uog.17435] [Medline: 28195383]

20. Infante F, Espada Vaquero M, Bignardi T, Lu C, Testa AC, Fauchon D, et al. Prediction of Tubal Ectopic Pregnancy Using Offline Analysis of 3-Dimensional Transvaginal Ultrasonographic Data Sets: An Interobserver and Diagnostic Accuracy Study. J Ultrasound Med 2018 Jun;37(6):1467-1472. [doi: 10.1002/jum.14489] [Medline: 29219200]
21. Chalouhi GE, Bernardi V, Gueneuc A, Houssin I, Stirnemann JJ, Ville Y. Evaluation of trainees' ability to perform obstetrical ultrasound using simulation: challenges and opportunities. Am J Obstet Gynecol 2016 Apr;214(4):525.e1-525.e8. [doi: 10.1016/j.ajog.2015.10.932] [Medline: 26546849]
22. Bultez T, Bernard J, Metzger U, Ville Y, Salomon LJ. Pilot Study of a Software-Supported Protocol for Second-Trimester Ultrasound Screening. J Ultrasound Med 2017 Nov;36(11):2279-2285. [doi: 10.1002/jum.14267] [Medline: 28600890]
23. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 2013;346:f657 [FREE Full text] [Medline: 23412440]
24. Yeo L, Romero R. Intelligent navigation to improve obstetrical sonography. Ultrasound Obstet Gynecol 2016 Apr;47(4):403-409 [FREE Full text] [doi: 10.1002/uog.12562] [Medline: 26525650]
25. Garcia M, Yeo L, Romero R, Haggerty D, Giardina I, Hassan SS, et al. Prospective evaluation of the fetal heart using Fetal Intelligent Navigation Echocardiography (FINE). Ultrasound Obstet Gynecol 2016 Apr;47(4):450-459 [FREE Full text] [doi: 10.1002/uog.15676] [Medline: 26278116]
26. Moratalla J, Pintoffl K, Minekawa R, Lachmann R, Wright D, Nicolaides KH. Semi-automated system for measurement of nuchal translucency thickness. Ultrasound Obstet Gynecol 2010 Oct;36(4):412-416 [FREE Full text] [doi: 10.1002/uog.7737] [Medline: 20617517]
27. Tutschek B, Pilu G. Pocket Brain, an interactive, web-based ultrasound atlas of normal and abnormal fetal brain development. Ultrasound Obstet Gynecol 2017 Dec;49(4):431-432 [FREE Full text] [doi: 10.1002/uog.15937] [Medline: 27071832]
28. Demner-Fushman D, Kohli MD, Rosenman MB, Shooshan SE, Rodriguez L, Antani S, et al. Preparing a collection of radiology examinations for distribution and retrieval. J Am Med Inform Assoc 2016 Mar;23(2):304-310 [FREE Full text] [doi: 10.1093/jamia/ocv080] [Medline: 26133894]
29. Kahn CE, Thao C. GoldMiner: a radiology image search engine. AJR Am J Roentgenol 2007 Jun;188(6):1475-1478. [doi: 10.2214/AJR.06.1740] [Medline: 17515364]
30. Friszer S, Dhombres F, Morel B, Zerah M, Jouannic JM, Garel C. Limited Dorsal Myeloschisis: A Diagnostic Pitfall in the Prenatal Ultrasound of Fetal Dysraphism. Fetal Diagn Ther 2017;41(2):136-144. [doi: 10.1159/000445995] [Medline: 27160821]
31. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res 2017 Dec 04;45(D1):D865-D876 [FREE Full text] [doi: 10.1093/nar/gkw1039] [Medline: 27899602]


## Abbreviations

**CDI:** color Doppler imaging
**CDSS:** clinical decision support system
**EP:** ectopic pregnancy
**OB/GYN:** obstetrics and gynecology
**TVS:** transvaginal sonography


*Edited by G Eysenbach; submitted 05.04.19; peer-reviewed by S Sabarguna, S Sarbadhikari; comments to author 08.05.19; revised version received 11.06.19; accepted 11.06.19; published 03.07.19.*

*Please cite as:*
*Dhombres F, Maurice P, Guilbaud L, Franchinard L, Dias B, Charlet J, Blondiaux E, Khoshnood B, Jurkovic D, Jauniaux E, Jouannic JM*
*A Novel Intelligent Scan Assistant System for Early Pregnancy Diagnosis by Ultrasound: Clinical Decision Support System Evaluation Study*
*J Med Internet Res 2019;21(7):e14286*
*URL: http://www.jmir.org/2019/7/e14286/*
*doi:10.2196/14286*
*PMID:31271152*

Original Paper

# Avatar-Based Patient Monitoring With Peripheral Vision: A Multicenter Comparative Eye-Tracking Study

Juliane Pfarr[1*], MD; Michael T Ganter[2], MD; Donat R Spahn[1], MD; Christoph B Noethiger[1], MD; David W Tscholl[1*], MD

[1]Institute of Anesthesiology, University and University Hospital Zurich, Zurich, Switzerland

[2]Institute of Anesthesiology Kantonsspital Winterthur, Winterthur, Switzerland

[*]these authors contributed equally

**Corresponding Author:**
David W Tscholl, MD
Institute of Anesthesiology
University and University Hospital Zurich
Raemistrasse 100
Zurich, 8091
Switzerland
Phone: 41 +41786636787
Email: david.tscholl@usz.ch

## Abstract

**Background:** Continuous patient monitoring has been described by the World Health Organization as extremely important and is widely used in anesthesia, intensive care medicine, and emergency medicine. However, current state-of-the-art number- and waveform-based monitoring does not ideally support human users in acquiring quick, confident interpretations with low cognitive effort, and there are additional problematic aspects such as alarm fatigue. We developed a visualization technology (Visual Patient), specifically designed to help caregivers gain situation awareness quickly, which presents vital sign information in the form of an animated avatar of the monitored patient. We suspected that because of the way it displays the information as large, colorful, moving graphic objects, caregivers might be able to perform patient monitoring using their peripheral vision, which may facilitate quicker detection of anomalies, independently of acoustic alarms.

**Objective:** In this study, we tested the hypothesis that avatar-based monitoring, when observed with peripheral vision only, increases the number of perceptible changes in patient status as well as caregivers' perceived diagnostic confidence compared with a high-fidelity simulation of conventional monitoring, when observed with peripheral vision only.

**Methods:** We conducted a multicenter comparative study with a within-participant design in which anesthesiologists with their peripheral field of vision looked at 2 patient-monitoring scenarios and tried to identify changes in patient status. To ensure the best possible experimental conditions, we used an eye tracker, which recorded the eye movements of the participants and confirmed that they only looked at the monitoring scenarios with their peripheral vision.

**Results:** Overall, 30 participants evaluated 18 different patient status changes with each technology (avatar and conventional patient monitoring). With conventional patient monitoring, participants could only detect those 3 changes in patient status that are associated with a change in the auditory pulse tone display, that is, tachycardia (faster beeping), bradycardia (slower beeping), and desaturation (lower pitch of beeping). With the avatar, the median number of detected vital sign changes quadrupled from 3 to 12 ($P$<.001) in scenario 1, and more than doubled from 3 to 8 ($P$<.001) in scenario 2. Median perceived diagnostic confidence was *confident* for both scenarios with the avatar and *unconfident* in scenario 1 ($P$<.001), and *very unconfident* in scenario 2 ($P$=.024) with conventional monitoring.

**Conclusions:** This study introduces the concept of peripheral vision monitoring. The test performed showed clearly that an avatar-based display is superior to a standard numeric display for peripheral vision. Avatar-based monitoring could potentially make much more of the patient monitoring information available to caregivers for longer time periods per case. Our results indicate that the optimal information transmission would consist of a combination of auditory and avatar-based monitoring.

***(J Med Internet Res 2019;21(7):e13041)*** doi:10.2196/13041

**KEYWORDS**

anesthesia; critical care; computers; diagnosis; patient monitoring; situation awareness; perception; vision

## Introduction

### Patient Monitoring Background

In its Guidelines for Safe Surgery, the World Health Organization describes continuous patient monitoring by an attentive and professionally trained caregiver as *extremely important* for perioperative safety [1]. Noninvasive standard monitoring offers an excellent risk-benefit ratio, as it is not dangerous for patients, yet through earlier, clearer detection of vital sign abnormalities than is possible by assessing clinical signs alone, it may prevent potential catastrophic complications, for example, brain damage [2-6]. Patient monitoring enjoys widespread acceptance among caregivers and professional associations in anesthesia, intensive care medicine, and emergency medicine [1,7]. With technological progress in sensor and computer technology, patient monitoring can be expected to increasingly be extended to areas where patients are currently not routinely monitored, thereby detecting vital sign anomalies even earlier than is now the case, for example, in general hospital wards [8].

### Introduction to Conventional Patient Monitoring

Human factor experts have long recognized that representation of vital sign data in the form of a multitude of numbers and waveforms in today's state-of-the-art monitors does not ideally support human users in arriving at a quick interpretation with a high degree of confidence and with a low cognitive effort [9]. Several characteristic aspects of conventional representation are responsible for this: (1) people can only read numbers one by one [10]; (2) the numbers displayed represent low-level data and, only indirectly, the relevant information [11]; (3) many of the numbers displayed have the same ranges, for example, pulse rate, blood pressure, oxygen saturation, and others can all be *95*; (4) people can only remember 7 digits plus or minus 2 at a time in their short-term memory [12]. The resulting need for piecemeal data acquisition, mental decoding, and subsequent interpretation of the meaning of the data requires much time and cognitive effort on the part of the caregiver to obtain adequate situation awareness of the patient's current condition. Situation awareness refers to the correct perception of a situation and its expected course [13]. It is an essential prerequisite for informed decision making, and research has identified situation awareness errors in up to 80% of adverse events [14,15]. Patient monitors, to mitigate some of their limitations, use audible and visual alarms to warn caregivers when vital signs diverge from their normal range. However, around 80% of issued alarms are false-positives that do not lead to a therapeutic consequence, leading to a *crying wolf phenomenon*, that is, caregivers experiencing alarm fatigue, with resulting failure to detect truly positive alarms [16]. In a recent study, 56% (14/25) of anesthesiologists agreed with the statement that problems with alarm settings make their work with patient monitors more difficult [17]. Studies investigating patient monitoring behavior have found that anesthesia providers look at patient monitors for only about 5% of the time during a procedure and that they tend to look less often in high-workload situations, when other tasks cause cognitive saturation [18,19].

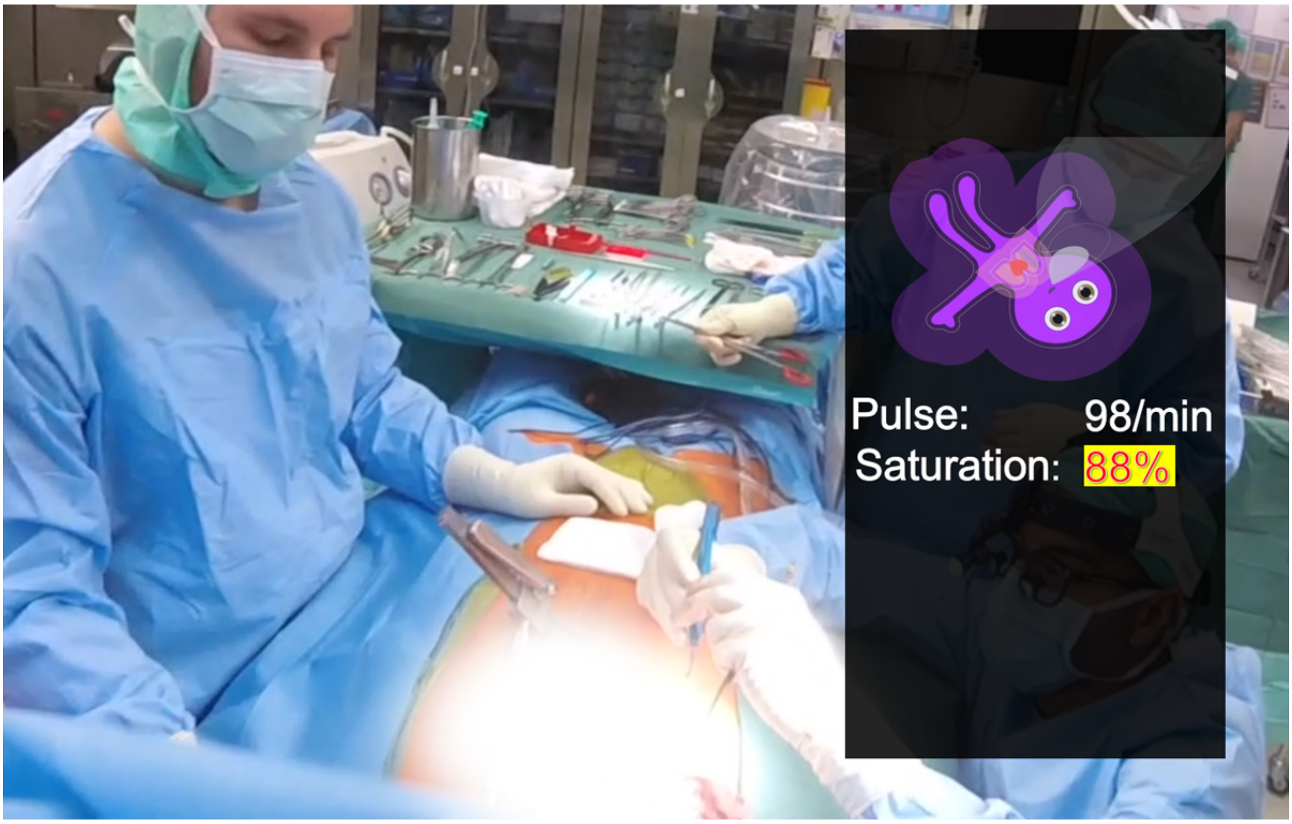
### Introduction to Avatar-Based Monitoring

One possible way to optimize the process of information transfer between patient monitors and caregivers may be to present the vital sign information in the form of graphical objects [9,20,21]. Applying principles of situation awareness design [22], we developed an avatar-based visualization technology (Visual Patient) that presents vital sign information as an animated avatar of the monitored patient and is specifically aimed at helping caregivers gain situation awareness quickly and with low cognitive effort. In a previous study [23], we found that compared with conventional technology monitoring the same short clinical scenarios, this technology increased both the number of correctly perceived vital signs and the diagnostic confidence reported by the participating anesthesia providers, while reducing perceived workload. Furthermore, users considered the technology intuitive, easy to learn, and helpful [24].

### Patient Monitoring With Peripheral Vision

We suspected that because of the way the avatar representation displays the information as large, colorful, moving graphical objects, caregivers might be able to perform patient monitoring using their peripheral vision. Conventional monitoring is particularly unsuitable for monitoring with peripheral vision because of the presentation of information in the form of numbers and figures as described above. To be able to read a number, a caregiver must fix their foveal or sharp vision directly on the number they intend to read. Foveal vision corresponds to the small central part of the retina in which a large number of cones are concentrated. Outside the area of foveal vision, color perception deteriorates and vision becomes blurry, rendering people unable to read glyphs with their peripheral vision [10,25,26]. Patient monitoring with peripheral vision could provide several theoretical advantages. It could increase the time per case that a caregiver has direct visual contact with the monitoring information, from approximately 5% of the time during which they observe the monitor with foveal vision to all the time they have the monitor in their peripheral visual field. Considering that the human binocular visual field encompasses approximately 214 arc degrees horizontally and 150 arc degrees vertically, the monitor is within the visual field virtually all of the time [27]. Peripheral vision monitoring may facilitate quicker detection of anomalies independently of acoustic alarms. Furthermore, the feeling of always keeping an eye on the situation may reduce caregiver stress levels, as uncertainty is a psychological stress factor [22,28]. Figure 1 shows an example of a potential future use of peripheral vision monitoring.

**Figure 1.** An example of a possible future application of peripheral vision monitoring in the form of an augmented reality application for patient monitoring, as Philips (Koninklijke Philips NV, Amsterdam, Netherlands) has tested on a Google (Alphabet Inc) Glass headset. If the reader looks at the center of the operating field in this photo, they can no longer read the numerical monitoring information, for example, saturation: 88%, however, they can still see that the avatar is purple and thus desaturated.




### Objective

In this study, we tested the hypothesis that avatar-based monitoring with peripheral vision increases the number of vital signs perceptible as well as perceived diagnostic confidence compared with conventional monitoring with peripheral vision.

## Methods

The Cantonal Ethics Committee in Zurich, Switzerland, reviewed the protocol of this study and issued a declaration of no objection (Business Administration System for Ethical Committees-Number 2017-00795 issued on October 23, 2017). All participants gave their written informed consent to the use of the data collected for scientific evaluation. The participants participated voluntarily in this study and received no financial compensation.

### Description of Visual Patient Technology

The version of the technology used in this study can display the 11 most commonly monitored vital signs: pulse rate, blood pressure, oxygen saturation, ST segment of the electrocardiogram, central venous pressure, respiratory rate, tidal volume, expiratory carbon dioxide concentration, body temperature, brain activity, and degree of neuromuscular relaxation.

We developed the avatar as a situation-awareness tool, analogous to the so-called synthetic vision technology in aviation, and according to the principles of situation-awareness design and logic [22,29]. The synthetic vision technology renders a virtual image of the environment from data measured by the aircraft, for example, altitude information, and global positioning system–referenced elevation data. To the pilot, the virtual image it creates looks identical to the view outside the window in perfect weather. This similarity is what makes it intuitively understandable and allows for the quick and uncomplicated perception of the flight situation. Visual Patient technology does the same, in this case, by creating a virtual image of the patient from vital sign data. Similar to synthetic vision technology, it presents the numerical data in a way that corresponds to the physical phenomena they engender in the patient. For example, low oxygen saturation is represented with cyanotic skin color because this is what hypoxia causes in a patient, and that is what caregivers expect.

This so-called direct presentation of information eliminates the necessity for a caregiver to calculate the relevant information mentally from lower level data, for example, is the patient hypoxic or not if oxygen saturation is 85% [11]. In addition to this direct presentation of information, the 2 other main characteristics of the avatar are the preprocessing of the data for each vital sign into the categories *too low*, *normal*, or *too high* and the presentation of the vital sign information in several visualizations at the same time. For example, caregivers can judge the respiratory rate by the respiratory rate of the avatar's lungs as well as the formation rate of the carbon dioxide cloud exhaled by the avatar.

These functions combined translate the multitude of numerical values into an animated model of the patient situation, which the caregiver can evaluate and remember at a glance. The translation of the vital signs into the avatar model takes place in real time from the monitoring data. We have described the validation and evaluation process of the avatar in detail in previous studies [23,24].

### Study Participants

For this study, we included anesthesia providers in 2 study centers. The University Hospital of Zurich, a University maximum care hospital with more than 30,000 anesthesia cases per year and the Cantonal Hospital of Winterthur, a regional teaching hospital with more than 10,000 cases per year. Both centers included balanced proportions of female and male participants as well as equal proportions of the different professional groups (senior and resident physicians and anesthesia nurses). To participate in the study, we freed the participants from their respective tasks during their regular working hours so that they could participate undisturbed by external influences.

### Study Procedure

We collected the data for this experiment as part of a session in which we also collected the data for 2 more experiments. Each of the participants sat in a quiet room of the University Hospital Zurich or the Cantonal Hospital Winterthur accompanied by a data collector, who guided the participant through the experiments. During the experiments, the participants sat in front of 2 computer screens. Initially, they watched an instructional video about Visual Patient technology and familiarized themselves with the layout of the conventional monitoring display. After the introduction and after they had completed a short personal information questionnaire (gender, age, and years of professional anesthesia experience), the experiments were conducted in sequence. In a pilot study, we discovered that the approximate duration of 1 data collection session would be about 1 hour and 15 min. With 2 short pauses between the 3 experiments, we considered this duration acceptable for the participants' ability to remain concentrated during the tests. The peripheral vision experiment conducted for this study was experiment number 3. We will report on the results of experiments 1 and 2 in separate papers. For all 3 experiments, we used an iPad-based (Apple Inc) data entry tool for data entry during the experiments [30].

### Peripheral Vision Experiment

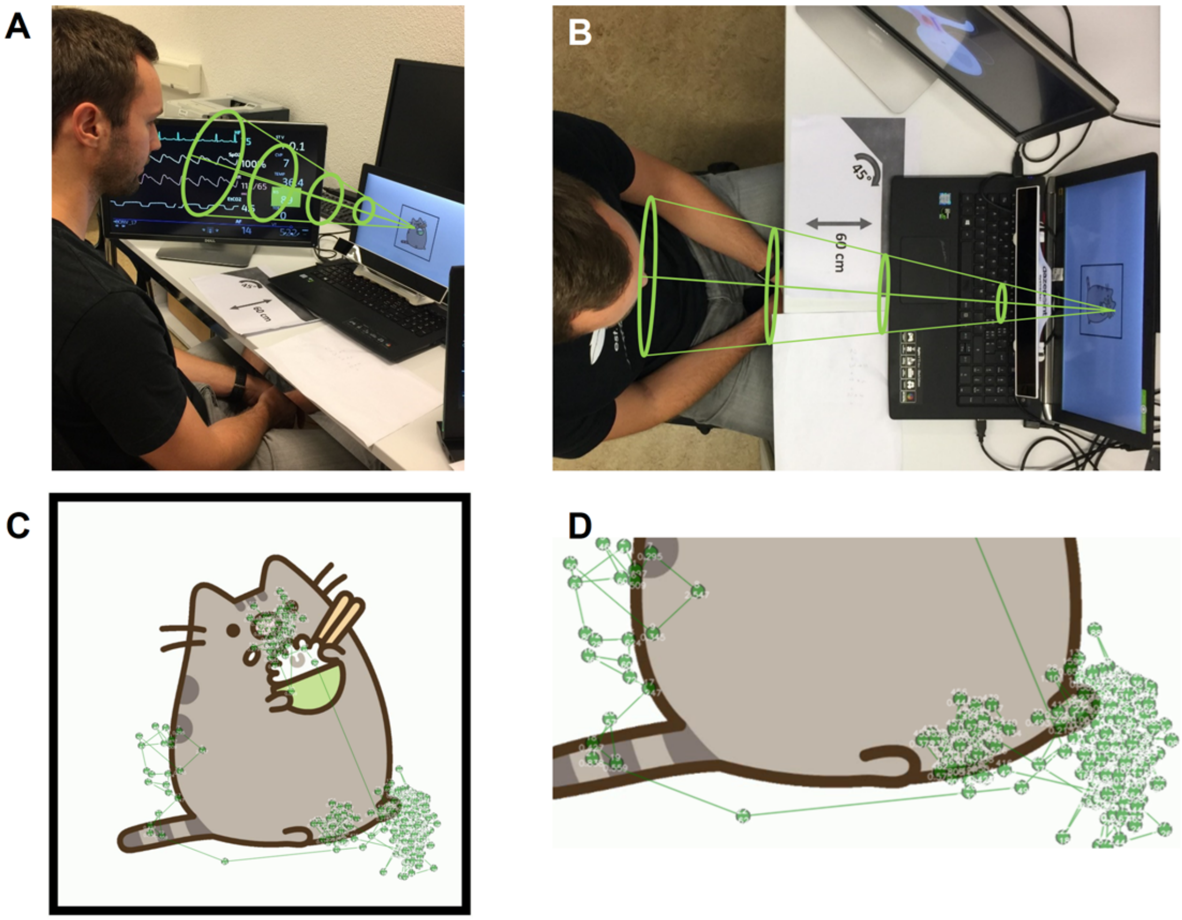
As the first step in the peripheral vision experiment, we positioned the participants at a distance of approximately 60 cm directly in front of a laptop screen. Then a stationary eye tracker (Gazepoint GP3 by Gazepoint) was calibrated to capture the foveal vision, that is, gaze plot, a sequence and durations of visual fixations, of the participants on the laptop screen directly in front of them.

For the peripheral vision scenario, we played the monitoring scenario evaluated by the participants on the second screen on the left side of the participant, which we placed at an angle of 45° to the visual axis of the participants and, thus, in their peripheral field of view. We instructed the participants never to look away from the central screen during the entire 8-min test. On this central screen, we showed a Microsoft (Microsoft Corp) PowerPoint presentation showing an animated graphic of a cat in an endless loop. This ensured that the foveal vision of the participants remained on the central screen and that they, therefore, could only see the scenario played on the second monitor with their peripheral field of vision. This method ensured that the volunteers really only looked at the scenario with their peripheral vision and that the data collected were, therefore, valid for this evaluation. Even if the participant's view were to wander to the left edge of the screen, the monitor at a 45° angle, that is, the monitor on which the monitoring scenario was running, would still be deep in the middle peripheral field of view, which extends from 30° to 60° from the point of sharpest vision.

We used this method based on research that showed that the human observer's vision is only sharp enough to read numbers or glyphs at an angle of 10° around the point of sharp vision. At a distance of 60 cm from the screen, the area of sharp vision is about the size of a fingernail, or more precisely, a circle about 2 cm in diameter. Outside of this small area, people only recognize blurred images and monochromatic colors [10,25]. In this study, we only included participants who did not look to the left of the monitor more than twice during the test. Two short glances at the monitor would, if both glances were successful and allowed for perception of 2 status changes, only account for a maximum of 5% of the participant's dataset, given the 36 evaluated status changes per dataset.

Figure 2 shows a picture of the experimental setup. A video showing a complete peripheral vision scenario is available (Multimedia Appendix 1). We recorded this video with a wide-angle camera, which we placed in such a way that it is possible for the reader to repeat the experiment by just looking at the cat and trying to identify the patient status changes on the screen to the left. The test works most realistically on a large screen and corresponds to our setup when the reader scales the video so that the central laptop screen has a diagonal of 15 inches.

**Figure 2.** (A and B) Study setup: A study participant sits in front of 2 computer monitors. An eye tracker records the participant's eye movements, which we used to confirm that the monitor on which the changes in patient condition were displayed was located in the peripheral field of view of the participant. The green funnel shows where the participant is looking and confirms that the monitor to the left remains in the peripheral visual field of the participant as long as they do not look away from the laptop screen in front. The base of the green cone corresponds to a radius of approximately 30° around the participant's point of sharpest vision. Everything outside the funnel lies in the participant's peripheral field of view. (C and D) The gaze plot data for 1 participant. Each point indicates a gaze fixation. A line links successive fixations.




### Scenarios

This experiment aimed to find out how many changes in patient status the participants could detect by peripheral vision with the 2 technologies (ie, avatar-based and conventional patient monitoring). For the 11 most important vital signs in today's clinical routine presented in our scenarios, there were 18 possible changes. Among them, 7 vital signs could become too high or too low (eg, blood pressure and pulse) and 4 vital signs could become abnormal only in 1 direction (eg, oxygen saturation). For the 2 technologies, we presented 36 scenarios in all. In these scenarios, all vitals remained normal for 5 seconds, after which one vital sign changed into the abnormal range. In the conventional scenarios, the changing vital signs were highlighted in yellow and an alarm tone sounded to allow for a high degree of realism. After each change, the data collector asked the test persons whether they had recognized which vital sign had just changed and, if yes, in which direction. The participants also indicated how confident they were that their assessment was correct. Participants were to choose from 0=very unconfident, 1=unconfident, 2=confident, and 3=very confident. In the videos, we showed the vital sign changes in randomized order, alternating avatar-based with conventional monitoring. To reduce the influence of the order in which the videos showed the vital sign changes, half of the participants evaluated a video in which we completely reversed the order of the vital sign changes compared with the first video.

### Outcome Measures

The primary objective of this study was to compare the performance of avatar-based monitoring with that of conventional patient monitoring in terms of the perception of patient status changes with peripheral vision. To quantify performance, we compared the number of recognized changes in vital signs with the 2 technologies. The higher the number of recognized vital sign changes, the more efficient the technology.

Secondary goals were to find out which vital signs the participants detected with the respective technologies and how confident they felt about the diagnoses they made.

### Statistical Analysis

As each participant evaluated the same vital sign changes using the 2 technologies, we used paired Student *t* test to check the differences for statistical significance. To compare subgroup data, we used Mann-Whitney U test, and for contingency tables, we used Fisher exact test, as appropriate.

### Sample Size

The sample size planning was based on the results of a pilot study and a post hoc sample size calculation for a paired *t* test. On the basis of these results, a sample size of 8 participants could demonstrate a difference in one of the 11 vital signs with a power of 80% at a significance level of 5%. In this calculation, we assumed that an improvement of a single perceived patient status change corresponds to the minimum clinically relevant difference.

For both scenarios, we expected significantly more than 8 participants and a higher difference than 1 patient status change between the technologies. Therefore, it was clear that with a total of 30 participants, the minimum requirements for a power of 80% at a significance level of 5% were exceeded.

## Results

### Study and Participant Characteristics

Table 1 shows the characteristics of this study and its participants in detail. Overall, 38 participants took part in the 2 study centers. Eye-tracking data were missing in 5 of the 38 participants because of technical recording problems. We excluded 3 more participants from the analysis because they looked to the left of the central monitor several times. In the end, we included data from 30 participants for evaluation. The 2 groups of participants at the Cantonal Hospital Winterthur and University Hospital Zurich were not significantly different in terms of gender, composition (professional groups), and anesthesia experience. The only difference was that participants from the Cantonal Hospital Winterthur more frequently belonged to a higher age group than did those from the University Hospital Zurich.

**Table 1.** Study and participant characteristics.

| Name of study center with number of participants. | Cantonal Hospital Winterthur (n=22) | University Hospital Zurich (n=16) | *P* value |
|---|---|---|---|
| Participants included in data analysis, n (%) | 17 (77) | 13 (81) | >.99[a] |
| Senior anesthesiologists, n (%) | 4 (25) | 5 (38) | .69[a] |
| Resident physicians, n (%) | 4 (25) | 2 (15) | .66[a] |
| Nurse anesthetists, n (%) | 8 (50) | 6 (46) | >.99[a] |
| **Number of female/male participants, n (%)** | | | |
| Female | 9 (56) | 8 (62) | .69[a] |
| Male | 7 (44) | 5 (38) | .69[a] |
| Age group of participants (years), median (IQR[b]) | 45-54 (25-34 to 45-54) | 25-34 (25-34 to 35-44) | . 05[c] |
| Anesthesia experience group of participants (years), median (IQR) | More than 10 (5-10 to >10) | 5 to 10 (1-5 to >10) | .32[c] |
| Duration of data collection (minutes), median (IQR) | 77 (70-86) | 76 (70-80) | .39[c] |
| Duration of peripheral vision experiment (minutes), median (IQR) | 13.5 (12-15) | 13 (12-15) | .43[c] |

[a]Fisher exact test.

[b]IQR: interquartile range.

[c]Mann-Whitney U test.

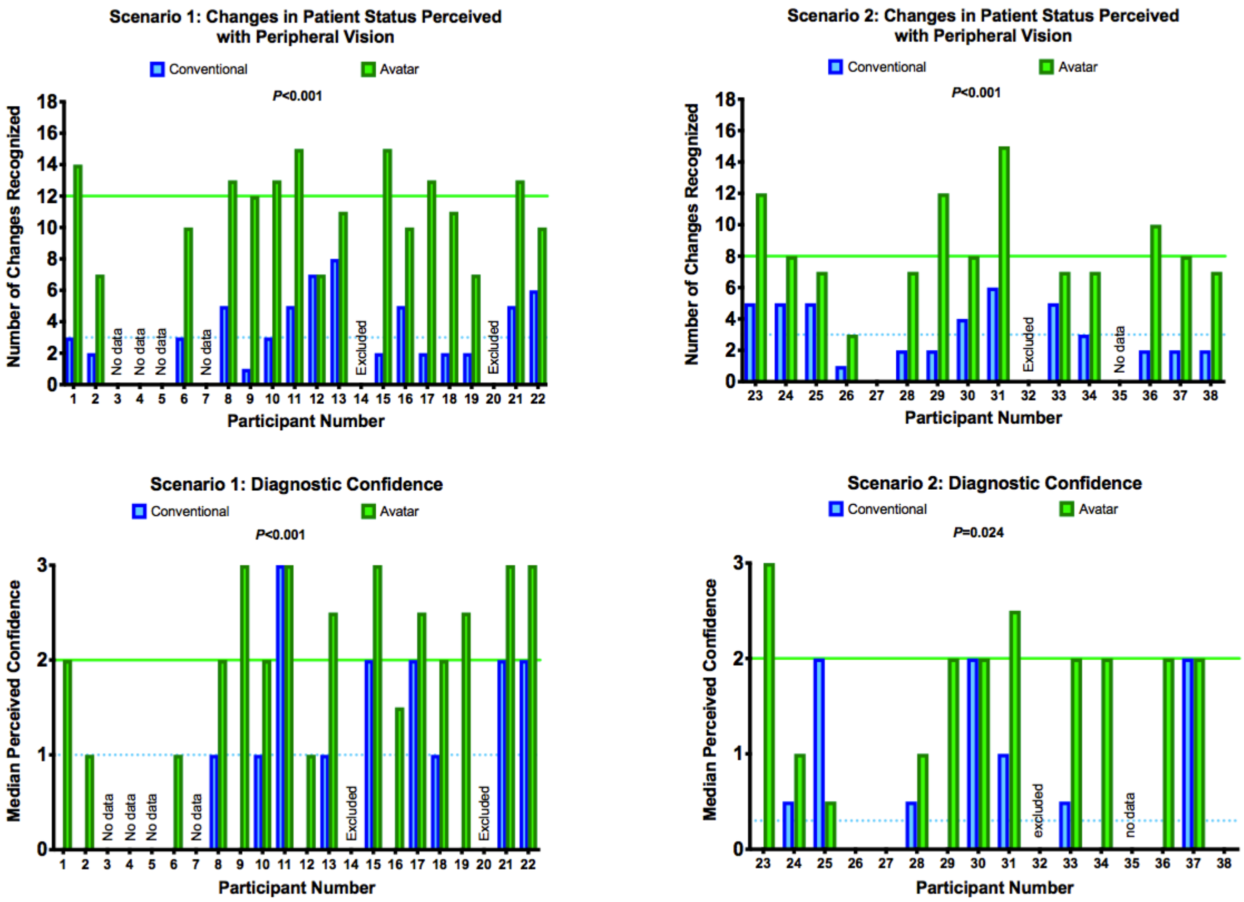
### Eye-Tracking Results

The eye-tracking data acquisition worked well. The success rate was 87% (33/38 participants). The most common reason for technical problems was thick eyeglass lenses, which did not allow for a successful calibration of the eye tracker. According to the study protocol, we excluded these data from the analysis. We provide the eye-tracking gaze plots of all individual participants (Multimedia Appendix 2).

### Primary Outcome

When the avatar was used, the number of changes in the patient's condition noticed with peripheral vision was higher. In scenario 1, it was higher by 9 vital signs, rising from a median (with interquartile range) of 3 (2-5) with the conventional monitor to 12 (10-13) with the avatar-based monitor (*P*<.001). In scenario 2, it was higher by 5 vital signs: increasing from 3 (2-5) with the conventional monitor to 8 (7-11) with the avatar-based monitor (*P*<.001). Figure 3 shows these results on individual participant level.

**Figure 3.** The results enabled 30 direct intraparticipant comparisons. All except 2 participants achieved a better performance with the avatar. The number of perceived changes in the patient's condition quadrupled in scenario 1 and more than doubled in scenario 2. Median perceived confidence: 0=very unconfident, 1=unconfident, 2=confident, and 3=very confident. Paired Student t tests showed statistical significance for all results.




Only 2 of 30 participants achieved the same result with conventional monitoring as with avatar-based patient monitoring. No participant performed better with conventional monitoring.

With conventional monitoring, only 2 changes in the patient's condition could be detected by more than half of the participants: *pulse too high* and *pulse too low*.

With the avatar, more than half of the participants recognized the following 8 out of 18 vital sign changes: (1) *pulse rate too high*, (2) *blood pressure too high*, (3) *saturation too low*, (4) *central venous pressure too high*, (5) *expiratory carbon dioxide concentration too high*, (6) *respiratory rate too high*, (7) *body temperature too high*, and (8) *body temperature too low*.

Only the vital sign change *pulse too low* was recognized by more participants with conventional monitoring than with avatar-based monitoring (Fisher exact test *P*<.001).

Table 2 shows exactly how many participants recognized each patient status change with each technology.

**Table 2.**

| Vital sign | Scenario 1 (n=16) | | | Scenario 2 (n=13) | | |
|---|---|---|---|---|---|---|
| | Conventional, n (%) | Avatar, n (%) | *P* value[a] | Conventional, n (%) | Avatar, n (%) | *P* value[a] |
| Pulse too high | 16 (100) | 16 (100) | >.99 | 13 (100) | 12 (92) | >.99 |
| Pulse too low | 8 (50) | 0 (0) | .002 | 9 (69) | 1 (8) | <.004 |
| Blood pressure too high | 0 (0) | 16 (100) | <.001 | 1 (8) | 13 (100) | <.001 |
| Blood pressure too low | 4 (25) | 4 (25) | >.99 | 0 (0) | 6 (46) | .010 |
| Saturation too low | 10 (63) | 15 (94) | .080 | 2 (15) | 9 (69) | .020 |
| Central venous pressure too high | 2 (13) | 15 (94) | <.001 | 0 (0) | 7 (54) | .005 |
| Central venous pressure too low | 0 (0) | 10 (63) | <.001 | 1 (8) | 2 (15) | >.99 |
| ST-Segment abnormal | 4 (25) | 7 (44) | .46 | 4 (31) | 2 (21) | .64 |
| Expiratory carbon dioxide concentration too high | 4 (25) | 16 (100) | <.001 | 3 (21) | 13 (100) | <.001 |
| Expiratory carbon dioxide concentration too low | 5 (31) | 10 (63) | .16 | 2 (15) | 1 (8) | >.99 |
| Respiratory rate too high | 1 (6) | 14 (88) | <.001 | 3 (23) | 11 (85) | .005 |
| Respiratory rate too low | 2 (13) | 2 (13) | >.99 | 2 (15) | 2 (15) | >.99 |
| Tidal volume too high | 2 (13) | 15 (94) | <.001 | 1 (8) | 6 (46) | .070 |
| Tidal volume too low | 2 (13) | 2 (13) | >.99 | 1 (8) | 1 (8) | >.99 |
| Brain activity high | 2 (13) | 9 (56) | .020 | 0 (0) | 5 (38) | .040 |
| Body temperature too high | 2 (13) | 16 (100) | <.001 | 0 (0) | 12 (92) | <.001 |
| Body temperature too low | 0 (0) | 9 (56) | <.001 | 0 (0) | 7 (54) | .005 |
| Neuromuscular relaxation high | 0 (0) | 5 (31) | .040 | 0 (0) | 2 (15) | .48 |

[a]Fisher exact test.

### Secondary Outcomes

The participants' perceived confidence in the correctness of their diagnoses reflected the higher number of perceived changes in the patient's condition. Only one of the 30 participants rated perceived confidence higher with conventional monitoring than with avatar-based monitoring. In scenario 1, median perceived confidence in the correctness of the diagnoses was *1*, that is, *unconfident* with conventional monitoring and *2*, that is, *confident* with avatar-based monitoring (*P*<.001). In scenario 2, this was *3*, that is, *very unconfident* with conventional monitoring and *2*, that is, *confident* with avatar-based monitoring (*P*<.001).

## Discussion

### Principal Findings

In this study, we found substantial differences between avatar-based and conventional patient monitoring with peripheral vision. In avatar-based monitoring, more than half of the participants in both scenarios detected the following 8 changes in patient status: (1) *pulse rate too high*, (2) *blood pressure too high*, (3) *oxygen saturation too low*, (4) *central venous pressure too high*, (5) *expiratory carbon dioxide concentration too high*, (6) *respiratory rate too high*, (7) *body temperature too high*, and (8) *body temperature too low*. In conventional patient monitoring, the only 2 changes that more than half of the participants in both scenarios detected were (1) *pulse rate too high* and (2) *pulse rate too low*. The pulse rate *too low* signal was the only one of the 18 total vital sign changes that was better detected with conventional monitoring (Table 2). Anesthesia providers are trained to detect a *too slow* pulse rate via the acoustic pulse tone because it implies serious a problem in a real patient. However, a nonpulsating avatar, as is the case for a low pulse rate, was not as well detected with peripheral vision. This finding suggests that optimal information transmission could be achieved with a combination of auditory and avatar-based monitoring; it also emphasized the potential benefits of further development of audio displays in patient monitoring, as also found previously [31].

### Potential Significance of Peripheral Vision for Patient Monitoring

Ford et al showed that anesthetists only looked directly at the screen of the patient monitor for about 5% of the time during anesthesia cases. This means that anesthesia providers spend much of their time at an angle where patient monitoring with peripheral vision would in theory be possible [18,19].

On the one hand, monitoring with peripheral vision has the theoretical advantage that a care provider can determine which vital signs lie outside the patient's normal range without having to look away from the patient or current tasks. This easing of workload would already be an advantage over today's industry standard monitoring devices and would become even more of an advantage with an augmented-reality head-mounted monitoring device in the future [32,33]. On the other hand, even during foveal viewing of the avatar, the information around the point of sharpest vision can still be perceived in parallel using peripheral vision. This method of information reception is not possible when reading a number from a conventional patient monitor interface, as humans can only read and process individual numbers in sequence [10,25]. Various sources have suggested that it is desirable for patient safety and operator well-being to design the exchange of information between an instrument and its human user as efficiently as possible [9,18,22]. A caregiver can only make the right decision for a patient if situation awareness is high. By definition, the concept of situation awareness encompasses 3 levels: (1) perception of elements in the environment within a volume of time and space, (2) understanding their meaning, and (3) projection of their status into the near future [13,22,34,35]. A lack of situational awareness prohibits sound decision making and is increasingly recognized as the cause of incidents and accidents in the medical field and in aviation [13-15,22,36]. Inadequate situation awareness constitutes a hole in the Swiss cheese model of Reason's theory of human error causation [37]. For patients who are connected to a patient monitor, the real-time and trend information from the screens and the acoustic displays are essential to the caregiver's situation awareness. Alarm fatigue and too frequent alarms are a major problem for anesthesia providers in their daily interaction with patient monitoring [17,38-40]. Avatar-based monitoring could theoretically provide a way to reduce audible alarms if, for example, an initial alarm was only visual and would only trigger an audible alarm after some time without a reaction.

The high ratings of the subjectively perceived certainty that their diagnoses were correct shows that the users had confidence in their assessment, despite the fact that they had only seen it with their peripheral vision and therefore hazily.

This study showed us that there are numerous potential advantages of patient monitoring by means of peripheral vision and that an animated patient avatar appears to be a good tool for evaluating the real-life usability of the concept of monitoring with peripheral vision. We plan to conduct further studies along this line.

### Limitations

We conducted this study as a computer-based laboratory study; therefore, it has limitations.

For instance, we have not yet tested the avatar in a real operating room and have not evaluated any clinical patient outcomes, such as clinical status of patients after surgery or adverse events. Although only a study conducted in a real-life environment or a high-fidelity simulator will ultimately allow for confident conclusions about the true benefits of avatar-based and peripheral vision monitoring, it is plausible that the large and reliable intraindividual improvements we observed would also manifest outside of the laboratory. In addition, we are at a very early stage of concept development, where it is crucial to identify potential theoretical benefits to determine whether clinical use is ultimately warranted.

In this study, we used a realistic simulation of a conventional monitor for conventional patient monitoring, including audio alarms and color highlighting of pathological values, such as the state-of-the-art devices that are currently in routine clinical use. It would be theoretically conceivable that further developments of these devices, which make pathological vital signs larger than *normal* vitals, would also be better readable with peripheral vision than today's devices. Nevertheless, even with critical numbers larger than the other numbers on a monitor, an avatar might have advantages because theoretically, although it is not yet tested, several vital signs could be readable simultaneously.

Another limitation is that we did not randomize the selection of participants for this study but that we recruited participants according to availability. However, we followed a plan made before the beginning of the study to include equal numbers of male and female participants and to balance participant numbers from the different occupational groups in the 2 centers. This standardization ensured that the groups were representative for all personnel groups and reduced the risk of sampling errors. Furthermore, the tendency of our anesthesia provider participants to look at the monitoring scenarios with foveal vision turned out to be low. Indeed, we only had to exclude 3 of 33 participants (10%) for looking at the peripheral monitor more than twice with foveal vision. All of the included participants looked at the peripheral monitor twice or less. Even if these 2 views of the monitor had allowed the perception of 2 status changes with foveal vision, this would have only affected 5% of the participants data, as each participant evaluated a total of 36 status changes. The low percentage of excluded participants and the eye-tracking method applied to confirm that the participants watched the scenarios with peripheral vision increase the validity of the study by reducing the risk of selection bias.

Finally, an inherent limitation of the avatar design that we would like to mention is the preprocessing of vital signs into categories, which causes a reduction of data accuracy. For example, pulse rate in the avatar can assume one of only 3 individual states, that is, *too low*, *normal*, or *too high*. Conversely, number and waveform monitoring can assume about 300 individual states between 0 and 300. Therefore, the avatar cannot replace routine monitoring, but it may serve as a supplement that explicitly aims at enhancing situation awareness.

### Conclusions

This study introduces the concept of peripheral vision monitoring. It provides empirical evidence that an avatar-based instrument can significantly improve the perception of patient status through peripheral vision. Further studies using the technology in real-life situations are necessary to show whether the benefits found can be translated into reality. This study represents a further building block in the literature on avatar-based monitoring by presenting and validating a hitherto unknown characteristic of the technology, namely, patient monitoring by means of peripheral vision.

## Acknowledgments

The authors are grateful to the study participants for their time and effort. Institute of Anesthesiology, University, and University Hospital Zurich provided institutional funding, Institute of Anesthesiology and Pain Therapy, Kantonsspital Winterthur provided institutional funding, University of Zurich provided proof-of-concept funding (UZ16/288POC), and University of Zurich provided DWT a career development grant.

## Authors' Contributions

DWT, JP, DRS, and CBN helped to design the study. DWT, JP, MTG, and CBN helped to collect the data. DWT, JP, MTG, DRS, and CBN helped to analyze the data. DWT, JP, MTG, DRS, and CBN helped to write the manuscript and approved the final version.

## Conflicts of Interest

The authors DWT, DRS, and CBN are in a joint development agreement with the monitoring manufacturer Philips Healthcare (Koninklijke Philips NV). Within the framework of this cooperation, a monitoring system based on an avatar will be developed. Within the framework of licensing the technology via the University, the authors DWT and CBN might receive royalties as designated inventors in the event of a successful product release.

## Multimedia Appendix 1

A video showing the complete peripheral vision test.

[MOV File, 175MB - jmir_v21i7e13041_app1.mov ]

## Multimedia Appendix 2

The eye-tracking gaze plots of all individual study participants.

[PDF File (Adobe PDF File), 6MB - jmir_v21i7e13041_app2.pdf ]

## References

1. WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives. Geneva: World Health Organization; 2009.
2. Mateer JR, Olson DW, Stueven HA, Aufderheide TP. Continuous pulse oximetry during emergency endotracheal intubation. Ann Emerg Med 1993 Apr;22(4):675-679. [doi: 10.1016/S0196-0644(05)81846-3] [Medline: 8457094]
3. Moller JT, Jensen PF, Johannessen NW, Espersen K. Hypoxaemia is reduced by pulse oximetry monitoring in the operating theatre and in the recovery room. Br J Anaesth 1992 Feb;68(2):146-150 [FREE Full text] [doi: 10.1093/bja/68.2.146] [Medline: 1540455]
4. Moller JT, Johannessen NW, Espersen K, Ravlo O, Pedersen BD, Jensen PF, et al. Randomized evaluation of pulse oximetry in 20,802 patients: II. Perioperative events and postoperative complications. Anesthesiology 1993 Mar;78(3):445-453. [Medline: 8457045]
5. Coté CJ, Rolf N, Liu LM, Goudsouzian NG, Ryan JF, Zaslavsky A, et al. A single-blind study of combined pulse oximetry and capnography in children. Anesthesiology 1991 Jun;74(6):980-987. [doi: 10.1016/0278-2391(92)90381-9] [Medline: 1904206]
6. Maheshwari K, Khanna S, Bajracharya GR, Makarova N, Riter Q, Raza S, et al. A randomized trial of continuous noninvasive blood pressure monitoring during noncardiac surgery. Anesth Analg 2018 Aug;127(2):424-431 [FREE Full text] [doi: 10.1213/ANE.0000000000003482] [Medline: 29916861]
7. Gelb AW, Morriss WW, Johnson W, Merry AF, Abayadeera A, Belîi N, International Standards for a Safe Practice of Anesthesia Workgroup. World Health Organization-World Federation of Societies of Anaesthesiologists (WHO-WFSA) international standards for a safe practice of anesthesia. Anesth Analg 2018 Jun;126(6):2047-2055. [doi: 10.1213/ANE.0000000000002927] [Medline: 29734240]
8. Michard F, Sessler DI. Ward monitoring 3.0. Br J Anaesth 2018 Nov;121(5):999-1001. [doi: 10.1016/j.bja.2018.07.032] [Medline: 30336872]
9. Drews FA, Westenskow DR. The right picture is worth a thousand numbers: data displays in anesthesia. Hum Factors 2006;48(1):59-71. [doi: 10.1518/001872006776412270] [Medline: 16696257]
10. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull 1998 Nov;124(3):372-422. [doi: 10.1037/0033-2909.124.3.372] [Medline: 9849112]
11. Degani A, Ames Research Center. On Organization of Information: Approach and Early Work. Moffett Field, California: National Aeronautics and Space Association Ames Research Center; May 2009.

12. Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev 1956 Mar;63(2):81-97. [doi: 10.1037/h0043158] [Medline: 13310704]
13. Schulz CM, Endsley MR, Kochs EF, Gelb AW, Wagner KJ. Situation awareness in anesthesia: concept and research. Anesthesiology 2013 Mar;118(3):729-742. [doi: 10.1097/ALN.0b013e318280a40f] [Medline: 23291626]
14. Schulz CM, Burden A, Posner KL, Mincer SL, Steadman R, Wagner KJ, et al. Frequency and type of situational awareness errors contributing to death and brain damage: a closed claims analysis. Anesthesiology 2017;127(2):326-337 [FREE Full text] [doi: 10.1097/ALN.0000000000001661] [Medline: 28459735]
15. Schulz CM, Krautheim V, Hackemann A, Kreuzer M, Kochs EF, Wagner KJ. Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. BMC Anesthesiol 2016 Jan 16;16:4 [FREE Full text] [doi: 10.1186/s12871-016-0172-7] [Medline: 26772179]
16. Schmid F, Goepfert MS, Kuhnt D, Eichhorn V, Diedrichs S, Reichenspurner H, et al. The wolf is crying in the operating room: patient monitor and anesthesia workstation alarming patterns during cardiac surgery. Anesth Analg 2011 Jan;112(1):78-83. [doi: 10.1213/ANE.0b013e3181fcc504] [Medline: 20966440]
17. Tscholl DW, Handschin L, Rössler J, Weiss M, Spahn DR, Nöthiger CB. It's not you, it's the design - common problems with patient monitoring reported by anesthesiologists: a mixed qualitative and quantitative study. BMC Anesthesiol 2019 May 28;19(1):87 [FREE Full text] [doi: 10.1186/s12871-019-0757-z] [Medline: 31138143]
18. Ford S, Birmingham E, King A, Lim J, Ansermino JM. At-a-glance monitoring: covert observations of anesthesiologists in the operating room. Anesth Analg 2010 Sep;111(3):653-658. [doi: 10.1213/ANE.0b013e3181e627d4] [Medline: 20581165]
19. Loeb RG. Monitor surveillance and vigilance of anesthesia residents. Anesthesiology 1994 Mar;80(3):527-533. [doi: 10.1097/00000542-199403000-00008] [Medline: 8141449]
20. Drews FA, Doig A. Evaluation of a configural vital signs display for intensive care unit nurses. Hum Factors 2014 May;56(3):569-580. [doi: 10.1177/0018720813499367] [Medline: 24930176]
21. Wachter SB, Johnson K, Albert R, Syroid N, Drews F, Westenskow D. The evaluation of a pulmonary display to detect adverse respiratory events using high resolution human simulator. J Am Med Inform Assoc 2006;13(6):635-642 [FREE Full text] [doi: 10.1197/jamia.M2123] [Medline: 16929038]
22. Endsley MR, Bolte B, Jones DG. Designing for Situation Awareness: An Approach to User-Centered Design. Boca Raton, Florida: CRC Press; 2011.
23. Tscholl DW, Handschin L, Neubauer P, Weiss M, Seifert B, Spahn DR, et al. Using an animated patient avatar to improve perception of vital sign information by anaesthesia professionals. Br J Anaesth 2018 Sep;121(3):662-671. [doi: 10.1016/j.bja.2018.04.024] [Medline: 30115265]
24. Tscholl DW, Weiss M, Handschin L, Spahn DR, Nöthiger CB. User perceptions of avatar-based patient monitoring: a mixed qualitative and quantitative study. BMC Anesthesiol 2018 Dec 11;18(1):188 [FREE Full text] [doi: 10.1186/s12871-018-0650-1] [Medline: 30537934]
25. Nielsen J, Pernice K. Eyetracking Web Usability. Berkeley, California: New Riders; 2010.
26. Hansen T, Pracejus L, Gegenfurtner KR. Color perception in the intermediate periphery of the visual field. J Vis 2009 Apr 30;9(4):26.1-2612. [doi: 10.1167/9.4.26] [Medline: 19757935]
27. Traquair HM. An Introduction to Clinical Perimetry. London: Kimpton; 1927.
28. Endsley MR, Jones WM. Situation Awareness Information Dominance & Information Warfare. Ohio: Wright-Patterson Air Force Base; 1997.
29. Wittgenstein L. Tractatus Logico-Philosophicus. London, UK: Routledge & Kegan Paul; 1922.
30. Tscholl DW, Weiss M, Spahn DR, Noethiger CB. How to conduct multimethod field studies in the operating room: the iPad combined with a survey app as a valid and reliable data collection tool. JMIR Res Protoc 2016 Jan 5;5(1):e4 [FREE Full text] [doi: 10.2196/resprot.4713] [Medline: 26732090]
31. Sanderson PM, Watson MO, Russell WJ, Jenkins S, Liu D, Green N, et al. Advanced auditory displays and head-mounted displays: advantages and disadvantages for monitoring by the distracted anesthesiologist. Anesth Analg 2008 Jun;106(6):1787-1797. [doi: 10.1213/ane.0b013e31817325cb] [Medline: 18499611]
32. Sanderson PM, Watson MO, Russell WJ. Advanced patient monitoring displays: tools for continuous informing. Anesth Analg 2005 Jul;101(1):161-168. [doi: 10.1213/01.ANE.0000154080.67496.AE] [Medline: 15976225]
33. Liu D, Jenkins SA, Sanderson PM, Watson MO, Leane T, Kruys A, et al. Monitoring with head-mounted displays: performance and safety in a full-scale simulator and part-task trainer. Anesth Analg 2009 Oct;109(4):1135-1146. [doi: 10.1213/ANE.0b013e3181b5a200] [Medline: 19762741]
34. Gaba DM, Howard SK, Small SD. Situation awareness in anesthesiology. Hum Factors 1995 Mar;37(1):20-31. [doi: 10.1518/001872095779049435] [Medline: 7790008]
35. Fioratou E, Flin R, Glavin R, Patey R. Beyond monitoring: distributed situation awareness in anaesthesia. Br J Anaesth 2010 Jul;105(1):83-90. [doi: 10.1093/bja/aeq137] [Medline: 20551029]
36. Endsley MR. Toward a theory of situation awareness in dynamic systems. Hum Factors 1995;37(1):32-64. [doi: 10.1518/001872095779049543]
37. Reason J. Human error: models and management. Br Med J 2000 Mar 18;320(7237):768-770 [FREE Full text] [doi: 10.1136/bmj.320.7237.768] [Medline: 10720363]

38. Ruskin KJ, Hueske-Kraus D. Alarm fatigue: impacts on patient safety. Curr Opin Anaesthesiol 2015;28(6):685-690. [doi: 10.1097/ACO.0000000000000260] [Medline: 26539788]
39. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
40. Srinivasa E, Mankoo J, Kerr C. An evidence-based approach to reducing cardiac telemetry alarm fatigue. Worldviews Evid Based Nurs 2017 Aug;14(4):265-273. [doi: 10.1111/wvn.12200] [Medline: 28432853]

*Edited by G Eysenbach; submitted 10.12.18; peer-reviewed by T Clavier, JJ Mira, C Fernández; comments to author 08.04.19; revised version received 03.06.19; accepted 12.06.19; published 17.07.19.*

*Please cite as:*
*Pfarr J, Ganter MT, Spahn DR, Noethiger CB, Tscholl DW*
*Avatar-Based Patient Monitoring With Peripheral Vision: A Multicenter Comparative Eye-Tracking Study*
*J Med Internet Res 2019;21(7):e13041*
*URL: http://www.jmir.org/2019/7/e13041/*
*doi:10.2196/13041*
*PMID:31317870*

Original Paper

# Differences in Perceptions of Health Information Between the Public and Health Care Professionals: Nonprobability Sampling Questionnaire Survey

Anat Gesser-Edelsburg[1], PhD; Nour Abed Elhadi Shahbari[1], BSc, MHA; Ricky Cohen[1], BSN, MHA; Adva Mir Halavi[1], BSN; Rana Hijazi[1], BPharm, MPH; Galit Paz-Yaakobovitch[1], BSc, RD, MPH; Yael Birman[1], BSc, RD, MPH

Health and Risk Communication Research Center and School of Public Health, Haifa, Israel

**Corresponding Author:**
Anat Gesser-Edelsburg, PhD
Health and Risk Communication Research Center and School of Public Health
University of Haifa
Haifa, 3498838
Israel
Phone: 972 544 243530
Fax: 972 3 6322143
Email: ageser@univ.haifa.ac.il

## Abstract

**Background:** In the new media age, the public searches for information both online and offline. Many studies have examined how the public reads and understands this information but very few investigate how people assess the quality of journalistic articles as opposed to information generated by health professionals.

**Objective:** The aim of this study was to examine how public health care workers (HCWs) and the general public seek, read, and understand health information and to investigate the criteria by which they assess the quality of journalistic articles.

**Methods:** A Web-based nonprobability sampling questionnaire survey was distributed to Israeli HCWs and members of the public via 3 social media outlets: Facebook, WhatsApp, and Instagram. A total of 979 respondents participated in the online survey via the Qualtrics XM platform.

**Results:** The findings indicate that HCWs find academic articles more reliable than do members of the general public (44.4% and 28.4%, respectively, *P*<.001). Within each group, we found disparities between the places where people search for information and the sources they consider reliable. HCWs consider academic articles to be the most reliable, yet these are not their main information sources. In addition, HCWs often use social networks to search for information (18.2%, *P*<.001), despite considering them very unreliable (only 2.2% found them reliable, *P*<.001). The same paradoxes were found among the general public, where 37.5% (*P*<.001) seek information via social networks yet only 8.4% (*P*<.001) find them reliable. Out of 6 quality criteria, 4 were important both to HCWs and to the general public.

**Conclusions:** In the new media age where information is accessible to all, the quality of articles about health is of critical importance. It is important that the criteria examined in this research become the norm in health writing for all stakeholders who write about health, whether they are professional journalists or citizen journalists writing in the new media.

***(J Med Internet Res 2019;21(7):e14105)*** doi:10.2196/14105

**KEYWORDS**

health information-seeking; reading and understanding; quality criteria for health journalists; Web-based and newspaper health information sources; journalistic articles; public healthcare workers and the general public

## Introduction

### Searching for Web-Based Health Information

In the last decade, the internet has become a powerful instrument for searching for health information because it provides the opportunity to access information from varied and diverse sources [1]. According to current estimates, over 4 billion people have access to the internet [2]. An analysis of global internet use indicates that Israelis are second among the top ten countries worldwide in time spent online among individuals age fifteen or above [3]. Israelis use the internet more than Americans and

Europeans [4] and spend the most time on the internet among global users [5].

As for science and health-related information, Israelis exhibit high levels of interest in science, with 62% of the public stating that knowing about science and technology in their everyday life is a necessity [6]. Moreover, polls have documented that health-related news is the most popular news topic among Israelis [7]. A recent study of otolaryngology patients found that Israelis turn to the internet as a source of health information significantly more than to books or newspapers [8].

Numerous studies have focused on factors that influence users in searching for health information. One such factor is health literacy [9-13]. Low health literacy is related to the limited ability to search, understand, and assess Web-based health information [14-16]. In contrast, high health literacy leads to more frequent Web-based searches for medical information [9,12].

Other variables found to influence the preference for internet sources include socioeconomic variables [13,17-19], cultural aspects [20,21], age variables [22-24], gender [25-28], and individual's level of self-efficacy [27,29,30].

One of the main consequences of searching for health information is its impact on actual behavior [31,32]. Some studies have indicated a connection between information seeking and actual behavioral improvement [33,34], with effective searching for Web-based health information leading to positive outcomes such as an improved understanding of medical conditions, an improved understanding of treatment options, informed decision making, and effective stress reduction [35-39]. On the contrary, Web-based health information can also contain misinformation and disinformation and can influence the behavior of people from various health fields [16,40-42].

### Traditional Information Sources Versus Web-Based Sources

Before the digital age, the general public depended heavily on health organizations and official sources for health information. The digital revolution provided new alternatives that enabled laypersons to rely on additional and alternative information sources and to self-manage their information [43-45]. The internet revolution has made it possible for people to take an active part in their medical care and manage their daily health needs [46]. This process has led to a shift in the perceived role of the public, from passive recipient to active consumer of health information [16,45,47-49].

Web-based media constitute an important source of health-related information [50,51], as well as a platform for discussing and sharing personal experiences, opinions, and concerns regarding illnesses and treatments [49,52-57]. Furthermore, the internet serves as a democratic, accessible, and interactive source of diverse information, thus enabling patients to make informed decisions [58,59]. Seeking and sharing Web-based information provides people with social support while enabling them to maintain their anonymity [60,61].

As a result of this revolution, health information no longer belongs exclusively to health professionals but rather is also accessible to the general public [22,48,62]. Thus, patients tend to bring information from Google to their appointments to discuss with their doctors so that the patient–physician discourse has changed from one-directional to two-directional communication [10,63-65]. During consultations with their doctors, patients seek to verify the information they obtained from other sources. After a doctor's appointment, some patients continue to search for information as a second opinion to verify the information they received from their doctor [18,62,66,67]. The reasons to continue searching information are as follows: the use of medical terminology in the physician–patient discourse impedes the patient's understanding [50] and the short duration of physician–patient encounters leads patients to seek other information sources to find answers to questions that remain open. These and other factors often lead the public to doubt the credibility of physicians and to consume medical information from Web-based sources [16,27,50].

In contrast, other studies show that despite the internet revolution, doctors are still considered the main source of reliable information [22,68-71]. Some studies indicate that searching for Web-based information can improve physician–patient communication [72]. Moreover, some patients do not see the internet as replacing the doctor but rather as another resource that can help them better understand medical recommendations [73]. Some studies indicate that consuming Web-based information also increases the public's reliance on medical professionals [74] for 2 main reasons: (1) the low reliability of the information on social networks [14,18,63,75-79] and (2) the low level of health literacy among the public, which makes it difficult for patients to understand and integrate health information and motivates them to turn to physicians as reliable sources [80-82].

In addition to doctors, before the digital revolution, the main agents through which the public obtained medical information were traditional media sources (eg, television and the press) [22,83-87]. Journalists who write about health acted as middlemen in communicating information to the public [16,42,88,89]. In today's age of new media, professional journalists constitute a new voice in the discourse, alongside citizens who also define themselves as journalists [16,42,90]. Nevertheless, health journalists still play an important role because they are considered to be professionals by the general public, which continues to read their Web-based articles [16,88,91]. For example, the study by Pew Research Center reported that 95% of all new information being disseminated via news media came from old media—especially newspapers [87]. According to Walsh-Childers et al (2018), because patients become increasingly involved in the direction of their health care:

> *...health journalism will likewise increase in salience for audiences as an education source. The current climate encourages a level of patient involvement in medical decision making that requires health care consumers to have a much better understanding of the benefits, harms, and costs of all options available to them.* [16]

The literature analyzing journalistic quality points to the problematic nature of transmitting reliable information. Numerous studies over the past decade have found many problems in the media coverage of medications and medical treatments. These problems primarily emerge in the tendency toward sensationalism and over-enthusiasm in describing medications and medical technologies by placing excessive emphasis on their benefits while ignoring or hardly mentioning their risks, side effects, or costs [92-98]. The literature also found that journalists rely increasingly on websites and press releases from the medical industry and health organizations, which can result in the public perceiving professional journalism as biased and lacking credibility [42,99-101].

In light of the aforementioned changes, it is important to examine how the public in general and health care workers (HCWs) in particular assess the quality of the journalistic articles they read.

### The Quality of Health Information

Retrieving Web-based information often leads to misinformation or disinformation [102]. The quality of the information offered on the internet varies, ranging from evidence-based scientific data quoting scientific research and clinical experiments to questionable information that could imperil the individual's health. Therefore, the challenge in searching for Web-based information and on social networks lies in the difficulty in finding sound, valid, and reliable information [103-105].

The proliferation of Web-based health information has led to a rise in the number of studies analyzing the quality of the published information. The first such study, published in 1997 by Impicciatore et al, evaluated the accuracy and integrity of the information on a website on fever management in children. According to the authors, out of 41 websites, only 4 provided full and accurate information about the subject [106]. This historic study provided a framework for subsequent studies of information quality. Accordingly, measurement tools were developed that usually included the following parameters: accuracy, completeness, readability, accountability, and technical criteria [103]. Additional specific parameters were added according to the subject of interest, for example, Eysenbach's 6 criteria [103], the WebMedQual Scale [107], and Godin's Quality Assurance Rating Tool for Internet Health Sites [108].

In recent years, these measurement tools and others have been used to assess information quality in different areas of medicine and health. All the studies point to the need to create reliable websites, improve information quality, improve access, increase oversight of Web-based medical information, and manufacture and distribute sound and customized materials [109-113]. Furthermore, websites have emerged that are dedicated to improving information quality for health writers. For example, the Health News Review website is an Australian-based website that aims to improve the public dialogue about health care by helping readers critically analyze health care news.

In general, most of the relevant literature in the new media age has focused on the reasons for seeking medical information, on diverse criteria that experts consider important to have on websites, and on the consequences of information seeking, whereas few studies have examined how the public reads and understands health information or what the public considers to be reliable.

This study sought to examine not only the ways in which the general public seeks information but also how the public understands this information and what information sources it sees as reliable. In addition, in view of the ongoing role of the press even in the new media era, the study also examines whether HCWs and the general public are capable of identifying quality criteria that influence behavioral intentions.

### Objectives

This research had 2 main objectives:

1. To examine how HCWs and the general public seek out, read, and understand health information.
2. To examine perceptions among HCWs and the general public regarding criteria for judging the quality of journalistic articles.

The specific objectives included the following: (1) to examine the differences between the general public and HCWs in how they seek and read information and how this new information influences their behavior and (2) to examine the criteria readers use to determine the quality of an article written by a journalist.

Comparing HCWs with the general public is based on the following rationale: As part of their daily work routine, HCWs are required to read and understand up-to-date information to be able to answer questions posed by the public. In addition, the public expects HCWs to answer questions about the information it encounters [10,16,63-65,72,114,115]. This raises questions regarding whether HCWs are able to discern the quality of journalistic information and whether they can refer the public to tools or criteria that can help in assessing the quality of articles appearing in the press.

The study is based on the following hypotheses:

1. The general public and HCWs will indicate that they will change their behavioral intentions after being exposed to health information.
2. The general public and HCWs will seek more health information from social networks than from scientific articles published in international journals.
3. The general public and HCWs will perceive scientific articles published in international journals as more reliable than information from social networks.
4. The general public and HCWs will be partially aware of the components that determine the quality of journalistic articles.
5. HCWs will perceive the criteria for determining the quality of health articles as more significant than will the general public.

## Methods

A survey was distributed to Israeli HCWs and members of the public via 3 social media outlets: Facebook, WhatsApp, and Google+. A total of 979 respondents participated in the online survey via the Qualtrics XM platform. The research was approved by the Faculty of Social Welfare and Health Sciences Ethics Committee for research with human subjects at the University of Haifa (Approval no. 266/18).

## Study Design

### *Sampling*

Our sample was designed using Qualtrics XM online survey software as it provided rapid and efficient distribution of an interactive online questionnaire (see Multimedia Appendix 1) to our research population (HCWs and the general public in Israel). We used the self-selection in Web survey method of nonprobability sampling [116] to recruit participants through posts on social networks asking the general public (over the age of 18 years) to answer the survey.

Development of the Questionnaire and Research Procedure

#### Stage 1: Building the Questionnaire

The questions were based on a literature review in the field of health information seeking and on HealthNewsReview.org, an Australian website (https://www.healthnewsreview.org/) designed to rank health articles according to quality criteria. The questionnaire consisted of 3 parts. The first part asked the participant for sociodemographic details.

The second part included questions about information searching and reliability attributed to health information sources (eg, “Where do you usually search for information?” and “Which of the aforementioned information sources (social networks, health organization websites, human sources, web-based newspapers, public healthcare workers) do you consider the most reliable?”). Respondents were also asked about health issues that concern them: “What main health area usually interests and concerns you; what type of health information do you search for and read about?” Respondents chose from a list of medical topics: nutrition, physical activity, illnesses, medications, vaccinations, alternative therapies, safety, environmental exposure, and other. Questions also focused on how the public reads health articles (eg, “Do you read the whole article or just parts of it?”) and the impact of the information on behavior change (eg, “If you encounter health information which seems important to you, to what extent would you change your behavior after learning about this?”). An example is information published by the World Health Organization indicating that processed meat raises the risk of cancer.

The third part of the questionnaire focused on what determines whether a health article is of high quality. Respondents were asked to rank the list of criteria they were given (see Table 1) on a 5-point Likert scale ranging from 1 (completely agree) to 5 (completely disagree).

**Table 1.** Importance of health information criteria to health care workers and to the general public.

| Criteria | Respondents[a] | |
|---|---|---|
| | Health workers | General public |
| The article also presents the drawbacks of the intervention. | 4.27 | 4.13 |
| The "tone" of the article is more scientific than marketing. | 4.18 | 3.96 |
| The article presents alternatives to medical intervention. | 4.17 | 4.05 |
| The article is based on a number of articles. | 4.16 | 3.86 |
| Details of the study. | 4.09 | 3.87 |
| The article cites results from an article from an academic journal. | 4.06 | 3.78 |
| Presentation of quantitative findings and not personal stories. | 4.04 | 3.84 |
| The article presents a scientific controversy in the field. | 4.01 | 3.80 |
| The article notes the availability and accessibility of treatment to the general public. | 3.97 | 3.94 |
| The article explains and simplifies professional concepts. | 3.96 | 4.03 |
| The article presents an opposing professional opinion. | 3.93 | 3.73 |
| The article presents existing conflicts of interest of the researchers. | 3.76 | 3.51 |
| The article presents information that has implications for policy. | 3.75 | 3.45 |
| The article presents a response by the regulator. | 3.72 | 3.56 |
| The article presents the findings even in the event that science indicates that there are no unequivocal answers. | 3.68 | 3.32 |

[a]Respondents were asked to rank the list of criteria they were given on a 5-point Likert scale ranging from 1 (completely agree) to 5 (completely disagree).

#### Stage 2: Criteria Validation Process

Before distributing the questionnaire to each population group, we analyzed the validity of the criteria using a focus group consisting of 25 students and 7 researchers at the School of Public Health (University of Haifa, Israel), who rated 3 health articles according to the criteria. We measured their overall percentage of agreement, as well as Krippendorff alpha (representing the level of agreement between coders beyond mere chance) for each answer [117]. Overall, for all the criteria, the between-coder agreement was satisfactory (alpha=.79; 92%).

### Stage 3: Pilot Survey for Content Validation

In the pilot, we distributed the questionnaire to 80 Arab and Jewish members of the general public and HCWs. The participants were asked to provide feedback on the questionnaire's content. Subsequently, we focused on adjusting the questions to make them more culturally sensitive. For example, we used the word regulator in Hebrew, but as that term is not used in Arabic, we modified the word to policy. Similarly, conflict of interest is not a familiar concept in Arabic, so in the interest of clarity, an example was provided. Furthermore, as the general public did not always understand the full meaning of a question, we added clarifying examples. For instance, for the item stating that "The article gives possible solutions for different medical issues", we gave examples of a possible solution (eg, making a lifestyle change instead of taking medicine to lower high blood pressure).

### Stage 4: Running the Study

To recruit as many participants as possible, we used intensive sampling in the first step and distributed the questionnaires via social media platforms (WhatsApp, Facebook, and Instagram). After this initial sampling, we continued to recruit participants through snowball sampling [118] to reach enough participants among HCWs by distributing the questionnaire via specialized HCW Web-based forums and by directly asking them to distribute the questionnaire to additional HCWs they knew.

At interim meetings during the survey, we monitored the social demographic variables and noted a lack of young men among the general public and the HCWs who responded to the survey. As our audience was a deliberate sample, we looked for ways to distribute the survey to more HCWs and turned to health forums. By means of diffusion, the survey was distributed from our inner circles to extended circles.

### Analysis

To check whether people intended to change their behavior after being exposed to health information, we used a chi-square test in which the answers are reduced to 3 levels (not at all or to a small extent, to a medium extent, and to a high to very high extent). Chi-square tests for independence were conducted to examine the differences between HCWs and others with regard to information seeking, sources of information, source reliability, and the manner in which the information was read. Wilcoxon Rank-Sum tests were used to examine the differences between HCWs and the public regarding behavioral change following exposure to information and the criteria for a high-quality article.

Regarding the quality criteria for articles, separate chi-square tests were conducted for each criterion to examine the differences between HCWs and others. To avoid the inflation of a type I error owing to multiple testing, adjusted P values were calculated using the false discovery rate method.

## Results

A total of 979 respondents participated in the survey. The vast majority of the respondents (96%) were below retirement age (<66 years) and female (76%), almost half were Jewish (49%) and somewhat fewer (42%) were Muslim. Table 2 depicts the respondents' sociodemographic and health status information.

**Table 2.** Sociodemographic and health status characteristics (n=979).

| Sociodemographic characteristics and category | n (%) |
|---|---|
| **Gender** | |
| Male | 232 (23.7) |
| Female | 747 (76.3) |
| **Age (years)** | |
| <29 | 363 (37.1) |
| 30-45 | 397 (40.6) |
| 46-65 | 177 (18.1) |
| 66+ | 42 (4.3) |
| **Ethnicity** | |
| Jewish | 481 (49.1) |
| Muslim | 410 (41.9) |
| Christian | 65 (6.6) |
| Druze | 12 (1.2) |
| Other | 11 (1.1) |
| **HCWs[a]** | |
| Yes | 363 (37.1) |
| No | 616 (62.9) |
| **Suffering from a chronic disease** | |
| Yes | 203 (20.7) |
| No | 776 (9.3) |
| **Child suffers from a chronic disease** | |
| Yes | 145 (14.8) |
| No | 834 (85.2) |

[a]HCWs: health care workers.

## Behavioral Intentions Following Exposure to Health Information: General Public Versus Health Care Workers

Respondents were asked about the extent to which they would change their behavior after receiving health information of personal importance: "If you encounter health information that seems important to you, to what extent would you change your behavior following your exposure to this information?" The research findings indicate that more than half the HCWs and more than half the respondents from the general public reported they would change their behavior to a large or very large extent (Table 3).

**Table 3.** Intention to change behavior after receiving health information of personal importance. Question: If you encounter health information that seems important to you, to what extent would you change your behavior following exposure to this information?

| Respondents | Intent to change | | | |
|---|---|---|---|---|
| | Not at all or to a low extent, n (%) | To a moderate extent, n (%) | To a large or very large extent, n (%) | Total |
| HCWs[a] | 20 (5.51) | 132 (36.36) | 211 (58.13) | 363 |
| GP[b] | 66 (10.71) | 220 (35.71) | 330 (53.57) | 616 |
| Total | 86 (8.78) | 352 (35.96) | 541 (55.26) | 979 |

[a]HCWs: health care workers.

[b]GP: general public.

### Seeking Health Information and Perceived Reliable Sources—the General Public Versus Health Care Workers

Table 4, Table 5 and Table 6 show where the general public and HCWs search for health information. In comparison with the general public, HCWs mainly search on health organization sites and in academic articles, as they consider academic articles more reliable. The general public seeks more information from social networks and Web-based newspapers and considers social networks, human resources, and HCWs to be more reliable.

**Table 4.** Seeking information and source reliability: comparison between health care workers and the general public.

| Sources for health information | Respondents, % | GP[a], % | HCWs[b], % | Chi-square (*df*) | *P* value | Adjusted *P* value[c] |
|---|---|---|---|---|---|---|
| **Social networks** | | | | | | |
| Where do you usually search for health information? | 30.34 | 37.50 | 18.18 | 40.33 (1) | <.001[d] | <.001[d] |
| Which source is most reliable in your opinion? | 6.13 | 8.44 | 2.20 | 15.45 (1) | <.001[d] | <.001[d] |
| **Health organizations** | | | | | | |
| Where do you usually search for health information? | 33.40 | 27.11 | 44.08 | 29.56 (1) | <.001[d] | <.001[d] |
| Which source is most reliable in your opinion? | 44.13 | 43.02 | 46.01 | 0.83 (1) | .36 | .36 |
| **Human sources** | | | | | | |
| Where do you usually search for health information? | 4.80 | 5.84 | 3.03 | 3.96 (1) | .05 | .06 |
| Which source is most reliable in your opinion? | 5.82 | 7.79 | 2.48 | 11.76 (1) | <.001[d] | <.001[d] |
| **Academic articles** | | | | | | |
| Where do you usually search for health information? | 17.06 | 12.82 | 24.24 | 21.05 (1) | <.001[d] | <.001[d] |
| Which source is most reliable in your opinion? | 34.32 | 28.41 | 44.35 | 25.76 (1) | <.001[d] | <.001[d] |
| **Public health care workers** | | | | | | |
| Where do you usually search for health information? | 5.52 | 5.68 | 5.23 | 0.09 (1) | .77 | .77 |
| Which source is most reliable in your opinion? | 8.27 | 10.71 | 4.13 | 13.04 (1) | <.001[d] | <.001[d] |
| **Web-based newspapers** | | | | | | |
| Where do you usually search for health information? | 8.89 | 11.04 | 5.23 | 9.51 (1) | .00[e] | .00[e] |
| Which source is most reliable in your opinion? | 1.33 | 1.62 | 0.83 | 1.11 (1) | .29 | .35 |

[a]GP: general public.

[b]HCWs: health care workers.

[c]False discovery rate.

[d]*P*<.001.

[e]*P*<.05.

**Table 5.** Primary information source and perception of reliability (percentage of health care workers).

| Information source | Source used to search for information | Most reliable source |
|---|---|---|
| Health organizations | 44% | 46% |
| Academic articles | 24% | 44% |
| Social networks | 18% | 2% |
| Public health workers | 5% | 4% |
| Human sources | 3% | 2% |
| Web-based newspapers | 5% | 1% |

**Table 6.** Primary information source and perception of reliability (percentage of general public).

| Information source | Source used to search for information | Most reliable source |
|---|---|---|
| Health organizations | 27% | 43% |
| Academic articles | 13% | 28% |
| Social networks | 38% | 8% |
| Public health workers | 6% | 11% |
| Human sources | 6% | 8% |
| Web-based newspapers | 11% | 2% |

## Health Care Workers Perceive the Criteria for Quality Health Articles as More Significant Than the General Public

The differences between HCWs and the general public in their perceptions of the importance of health information quality criteria are statistically significant, with the exception of 2 criteria: the article explains and simplifies professional concepts and the article notes the availability and accessibility of treatment to the general public. Inclusion of the criteria in the articles is more important for HCWs than for the general public (Table 1).

The comparison between the criteria rankings of the HCWs and those of the general public shown in Table 1 indicates that both groups ranked the following criteria at the top of the list: intervention drawbacks; tone more scientific than commercial; offers alternatives to medical interventions; based on several articles; and details of the study. Among the HCWs, the importance of citing results from academic articles was next on the list, whereas the general public ranked this criterion in the tenth place, instead ranking presentation of quantitative findings and not personal stories in the sixth place. Both the HCWs and general public ranked conflict of interest at the bottom of the chart (not shown in Table 7).

**Table 7.** Ranking of top 6 health information criteria among health care workers versus the general public.

| Group and criteria | 5-point Likert scale ranging from 1 (completely agree) to 5 (completely disagree) | *P* value[a] |
|---|---|---|
| **HCWs[b]** | | |
| Drawbacks of the intervention | 4.27 | .01[b] |
| Tone more scientific than commercial | 4.18 | .00[c] |
| Alternatives to medical interventions | 4.17 | .04[c] |
| Based on several articles | 4.16 | <.0001[d] |
| Details of the study | 4.09 | .00[c] |
| Cites results from academic articles | 4.06 | <.0001[d] |
| **General public** | | |
| Drawbacks of the intervention | 4.13 | .00[c] |
| Alternatives to medical interventions | 4.05 | .04[c] |
| Tone more scientific than commercial | 3.96 | .00[c] |
| Details of the study | 3.87 | .00[c] |
| Based on several articles | 3.86 | <.0001[d] |
| Presents quantitative findings and not personal stories | 3.84 | .00[c] |

[a]Wilcoxon Rank-sum Test

[b]HCWs: health care workers.

[c]*P*<.05.

[d]*P*<.001.

## Discussion

### Principal Findings

The new media age has changed the way people seek and consume health information [1,46]. The purpose of this study was to investigate not only how people search for Web-based and newspaper health information but also how they read and understand this information and what criteria they use to assess the quality of journalistic articles. It is important to examine how people read, understand, and assess the quality of journalistic health articles because health information can influence the way people shape their healthy lifestyles [33,34,119,120].

The findings of this study confirm the importance of this examination. When participants were asked whether they intended to change their behavior after being exposed to health information, more than 30% responded that they would make moderate changes to their behavior and 50% responded that they would make extensive changes.

The findings indicate that HCWs focus their search for health information on health organization sites and in academic articles, whereas the general public tends to search more on social networks and Web-based newspapers. This finding can be explained by the HCWs' professional context [121]. Public HCWs are accustomed to interacting with the health system on a daily basis and naturally search more on health organization websites [122,123]. Similarly, it is reasonable to assume that in the course of their work, professionals are more likely to use academic articles than the general public [124].

Moreover, the difference in choice of information sources between HCWs and the general public can also be explained by the level of health literacy. It is reasonable to assume that HCWs have a higher level of health literacy and are more capable of processing and understanding complex medical information than the general public, leading them to place more trust in scientific sources than in social networks or information available on the internet [12,125].

Furthermore, when participants were asked what they consider to be a reliable source, HCWs found academic articles more reliable than did the general public, which found social networks [55,58,59], human sources, and HCWs to be more reliable [68,74].

As for information seeking, HCWs found academic articles based on scientific facts to be more reliable than information from social networks. Moreover, they considered information based on scientific evidence from academic journals to be more reliable and to have a more scientific than commercial tone. Thus, differences between HCWs and the general public can be explained based on the training that HCWs undergo.

Nevertheless, when we examined each group separately, differences emerged within each group. Even though HCWs indicated that academic articles are the most reliable, they tended to search more for information on social networks despite considering them very unreliable [126]. We found a similar discrepancy among the general public, which considered health organizations and academic articles to be very reliable or reliable sources yet used them infrequently to search for health information, preferring social networks, which they considered unreliable. The research findings confirm our hypothesis that both HCWs and the general public search for health information on social networks more than they do in sources they consider reliable.

This discrepancy between perceptions and actual behavior is in line with studies indicating that there are situational factors whose influence is stronger than mere attitudes. According to Wicker (1969) [127], even though participants believe that health organizations and academic articles are more reliable sources than social networks, in practice, most of them search in sources they consider to be less reliable.

The following 3 explanations attempt to answer why the public seeks information from social networks more than from other sources. First, health organizations do not provide responses to the public's questions. That is, they conduct a monologue rather than a dialogue, leading the public to seek more information on social networks. The world of social media has generated a radical transformation in the relationship between government organizations and the public. Social media have changed the monologue to a dialogue in which anyone with information and access to communications technology can be a content creator and communicator [128]. Over the past decade, leading international health authorities, health ministries, and local governments have invested financial and human resources to narrow the gaps between the authorities and the public, thus increasing the authorities' presence on social media. Despite this impressive transformation, the use of social media by organizations is still in its infancy. Although the literature indicates that health authorities use social media, it also shows that this use is still very limited, as these tools serve primarily for mass information dissemination (similar to traditional mass media) instead of for 2-way communication [129-132].

Second, as health organizations do not exercise complete transparency in conveying information, the public turns to social networks to fill in the missing information. It is important to note that alongside disinformation deliberately conveyed by stakeholders, most of the discourse on social media stems from people's desire to obtain additional information, which is sometimes not fully conveyed by the health organizations [43].

Third, among the general public as well as among HCWs, decision making on health matters entails a combination between the automatic emotional system and the rational system. Thus, it is no wonder that despite being aware that social networks are likely to contain misinformation or disinformation, people continue to seek information there [133]. Studies of the health behavior of public HCWs found that they shared the same concerns and barriers as the general public [43,134-137]. Neither public HCWs nor the general public rely only on analytical or evidence-based information (academic articles) when searching for health information, but also seek information based on experience and emotions, both of which are found mostly on the social networks.

In summary, we proposed several possible explanations for the discrepancy we found between what the public and HCWs believe to be reliable and where they actually search for information in practice.

As for the research findings about what criteria the general public uses to judge the quality of journalistic articles compared with the criteria used by public HCWs, statistically significant differences were found between the importance of the criteria (except for 2), indicating that HCWs attributed more importance to the criteria than the public.

In addition, when we examined the 6 criteria that were most important to the general public and to HCWs, we found that 4 were important for both groups: drawbacks of the intervention, a tone that is more scientific than commercial, alternatives to medical interventions, and details of the study. These criteria indicate that the public values providing full information about the negative impacts or limitations of medical interventions as well as existing alternatives, information often absent from the media coverage. Studies have found that media coverage of medications and medical treatments is problematic, primarily in its tendency to provide sensational and an overly enthusiastic coverage of drugs and medical technologies and to emphasize the benefits excessively while ignoring or hardly mentioning the risks, side effects, and costs [93,94,96,97,98] or the limitations of scientific studies advocating the efficacy of these drugs.

In addition to the 4 aforesaid criteria cited both by public HCWs and by the general public as indications of information quality, 2 specific criteria emerged as important to the public. One of them is that the article should mention treatment availability and accessibility. This finding can be explained by the public's wish to know whether the treatment or medication mentioned in the article is accessible to it. For marketing reasons, press reports often mention medications and interventions that are not accessible to the public [92]. A second criterion valued by the public is that the article should simplify professional concepts. The importance of communicating professional information in understandable and clear language is a basic principle cited in the health communication and risk communication literature [138,139]. The mental models approach [140] also emphasizes the importance of understanding and simplifying professional concepts for the general public. Conversely, a criterion the public did not consider important was citing academic sources.

The findings also indicate that both HCWs and the general public ranked conflict of interest at the bottom of the list. Studies indicate that for years journalists have relied on information provided to them by organizations and the pharmaceutical industry rather than looking for quotes from academic sources themselves [99,100]. Several scholars have warned that journalists often fail to disclose the funding sources supporting the research, the investigators' financial conflicts of interests, and all the sources interviewed [95,141-143]. Owing to such potential conflicts of interest, reporting a study's limitations, funding sources, and financial ties is of great importance [144]. The Statement of Principles of the US Association of Health Care Journalists calls on journalists to disclose relevant conflicts of interest in their sources as a routine part of their work [145]. Yet, it seems that more often than not, journalists do not report such conflicts of interest [97,98,145].

The public may have become used to reading information in such a way that it does not look for citations from academic sources but relies on the author's integration or summary. In addition, the lack of discussion about the importance of exposing conflicts of interest leads both public HCWs and the general public not to attach adequate importance to this issue. The importance of including scientific articles and assessing their quality and the importance of disclosing conflicts of interest are criteria whose absence can produce misinformation, partial information, or disinformation that affect the public's decision making.

### Limitations

This study is not a representative sample of the general population of Israel. It used nonprobability sampling and measuring and was therefore vulnerable to selection bias from the outset.

Furthermore, in this study, we did not check the impact of several variables that might affect health-information searching behavior, both of the general public and of health workers, such as age, gender, personal relevance, level of health literacy, and the individual's reasons for the health information search. Also, as in any study checking behavioral intentions and actual behavior, this study is vulnerable to information bias as the result of biased reporting by the respondents. Our overall goal was to reach the specific target audiences of the general public versus HCWs and compare them, even though it was not a representative sample of those 2 populations.

We took a number of steps to minimize sampling bias: (1) we used 3 different media channels, thus increasing the chances for randomization in this sample [116]; (2) we monitored the data once a week to insure sufficient professional and ethnic representation among the participants. For example, when we discovered that there was an insufficient number of HCWs, we posted on more medical forums. When we noticed there were not enough participants from the Arab sector, we appealed specifically to this population group and thus broadened the sample; and (3) we used snowball sampling according to which each participant gave the questionnaire to someone else from their group, enabling us to reach more people from the required population groups. As our study is based on a small subpopulation of HCWs, the choice of the snowball sampling technique seemed to be more appropriate than convenience sampling. In addition, the descriptive statistics suggest that we were able to achieve a diverse sample based on sociodemographic variables.

### Conclusions

The study findings point to disparities both among HCWs and among the general public in their information-seeking behavior and their evaluations of the reliability of the sources searched. To reduce these gaps, health organizations must provide attractive materials, make academic articles accessible, and improve their dialogue with the public. In addition, in the technological age, where information is accessible to all, the quality of articles about health is critically important. Making the criteria cited in this research the norm in health writing is important for all stakeholders who write about health, whether they are professional journalists or citizen journalists in the new media.


## Acknowledgments

The authors wish to thank Ofir Raichman for his comments and suggestions at the early stages of the research and Dr Debby Mir for reading and editing the final version of manuscript.


## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Questionnaire.

[DOCX File, 33KB - jmir_v21i6e14105_app1.docx ]

## References


1. Galarce EM, Ramanadhan S, Weeks J, Schneider EC, Gray SW, Viswanath K. Class, race, ethnicity and information needs in post-treatment cancer patients. Patient Educ Couns 2011 Dec;85(3):432-439. [doi: 10.1016/j.pec.2011.01.030] [Medline: 21354761]
2. Internet World Stats. Internet Usage Statistics. The Internet Big Picture. World Internet Users and 2019 Population Stats URL: https://www.internetworldstats.com/stats.htm [accessed 2019-06-04] [WebCite Cache ID 78sdvcoJQ]
3. Israelashvili M, Kim T, Bukobza G. Adolescents' over-use of the cyber world--Internet addiction or identity exploration? J Adolesc 2012 Apr;35(2):417-424. [doi: 10.1016/j.adolescence.2011.07.015] [Medline: 21803411]
4. Gordon G, Rosenbloom T, Greifman Y, Zwillenberg P. The Boston Consulting Group. 2014 Jul. Digital Bloom. How Israeli consumers value Online media URL: https://www.bcg.com/about/people/experts/Tal-Rosenbloom.aspx [accessed 2019-06-04] [WebCite Cache ID 78seirjeG]
5. Epstein M. Haaretz. 2013 May 23. Israelis 'like' Facebook more than any other nation URL: https://www.haaretz.com/israel-news/business/.premium-israelis-like-facebook-more-than-other-nations-1.5268793 [accessed 2019-06-04] [WebCite Cache ID 78sfaviW0]
6. Dgani A, Dgani R. [The Seventh Eye Media and press criticism]. 2015 Apr. Perceptions and attitudes of the Israeli public science and space technology URL: https://cdn.the7eye.org.il/uploads/2015/06/%D7%A1%D7%A7%D7%A8-%D7%AA%D7%A4%D7%99%D7%A1%D7%95%D7%AA.pptx [accessed 2019-06-04] [WebCite Cache ID 78sg9n0Y2]
7. Israel Internet Association. 2009. Using sources of information on health and medicine in Israel URL: https://www.isoc.org.il/wp-content/uploads/2012/07/searchinghealthonline.pdf [accessed 2019-06-04] [WebCite Cache ID 78sgTAsKe]
8. Segal N, Gluck O, Sternbach D, Plakht Y, Segal A, Bachner YG. What are the sources for health information among patients attending otolaryngology outpatient clinics in Israel? A questionnaire survey. Br J Med Med Res 2015;5(6):775-779 [FREE Full text] [doi: 10.9734/BJMMR/2015/13506]
9. Britt RK, Hatten KN. Need for cognition and electronic health literacy and subsequent information seeking behaviors among university undergraduate students. SAGE Open 2013 Oct 29;3(4):1-10 [FREE Full text] [doi: 10.1177/2158244013508957]
10. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer Use of. J Med Internet Res 2015 Dec 29;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
11. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
12. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
13. Oh YS, Choi EY, Kim YS. Predictors of smartphone uses for health information seeking in the Korean elderly. Social work in public health. Soc Work Public Health 2018;33(1):43-54. [doi: 10.1080/19371918.2017.1391150] [Medline: 29257932]
14. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 7;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
15. Meppelink CS, Smit EG, Diviani N, Van Weert JC. Health literacy and online health information processing: unraveling the underlying mechanisms. J Health Commun 2016;21(sup2):109-120. [doi: 10.1080/10810730.2016.1193920] [Medline: 27668318]

16. Walsh-Childers K, Braddock J, Rabaza C, Schwitzer G. One step forward, one step back: changes in news coverage of medical interventions. Health Commun 2018 Dec;33(2):174-187. [doi: 10.1080/10410236.2016.1250706] [Medline: 27983868]
17. Dean CA, Geneus CJ, Rice S, Johns M, Quasie-Woode D, Broom K, et al. Assessing the significance of health information seeking in chronic condition management. Patient Educ Couns 2017 Aug;100(8):1519-1526. [doi: 10.1016/j.pec.2017.03.012] [Medline: 28320559]
18. Perez SL, Kravitz RL, Bell RA, Chan MS, Paterniti DA. Characterizing internet health information seeking strategies by socioeconomic status: a mixed methods approach. BMC Med Inform Decis Mak 2016 Dec 9;16:107 [FREE Full text] [doi: 10.1186/s12911-016-0344-x] [Medline: 27506607]
19. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
20. Dutta MJ. Communicating about culture and health: theorizing culture-centered and cultural sensitivity approaches. Commun Theory 2007 Aug;17(3):304-328 [FREE Full text] [doi: 10.1111/j.1468-2885.2007.00297.x]
21. Kim W, Kreps GL, Shin CN. The role of social support and social networks in health information-seeking behavior among Korean Americans: a qualitative study. Int J Equity Health 2015 Apr 28;14:40 [FREE Full text] [doi: 10.1186/s12939-015-0169-8] [Medline: 25927546]
22. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1-11 [FREE Full text] [doi: 10.1080/23311886.2017.1302785]
23. Miller Soederberg LM, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. J Aging Health 2012 Apr;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
24. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011 Feb 16;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
25. Ahadzadeh AS, Pahlevan Sharif S, Sim Ong F. Online health information seeking among women: the moderating role of health consciousness. Online Inf Rev 2018 Feb 12;42(1):58-72 [FREE Full text] [doi: 10.1108/OIR-02-2016-0066]
26. Altizer KP, Grzywacz JG, Quandt SA, Bell R, Arcury TA. A qualitative analysis of how elders seek and disseminate health information. Gerontol Geriatr Educ 2014;35(4):337-353 [FREE Full text] [doi: 10.1080/02701960.2013.844693] [Medline: 24188253]
27. Jaafar NI, Ainin S, Yeong MW. Why bother about health? A study on the factors that influence health information seeking behaviour among Malaysian healthcare consumers. Int J Med Inform 2017 Dec;104:38-44. [doi: 10.1016/j.ijmedinf.2017.05.002] [Medline: 28599815]
28. Manierre MJ. Gaps in knowledge: tracking and explaining gender differences in health information seeking. Soc Sci Med 2015 Mar;128:151-158. [doi: 10.1016/j.socscimed.2015.01.028] [Medline: 25618604]
29. Johnson JD, Case DO. Health Information Seeking, Volume 4. New York: Peter Lang Publishing Inc; 2012.
30. Lagoe C, Atkin D. Health anxiety in the digital age: an exploration of psychological determinants of online health information seeking. Comput Hum Behav 2015;52:484-491 [FREE Full text] [doi: 10.1016/j.chb.2015.06.003]
31. Johnson JD. Cancer-Related Information Seeking. Cresskill, NJ: Hampton Press; 1997.
32. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019 [FREE Full text] [doi: 10.1177/1049732307305199] [Medline: 17928475]
33. Lee CJ, Zhao X, Pena-y-Lillo M. Theorizing the pathways from seeking and scanning to mammography screening. Health Commun 2016;31(1):117-128. [doi: 10.1080/10410236.2014.942769] [Medline: 26086195]
34. Moldovan-Johnson M, Martinez L, Lewis N, Freres D, Hornik RC. The role of patient-clinician information engagement and information seeking from nonmedical channels in fruit and vegetable intake among cancer patients. J Health Commun 2014 Dec;19(12):1359-1376 [FREE Full text] [doi: 10.1080/10810730.2014.906521] [Medline: 24875456]
35. Huang GJ, Penson DF. Internet health resources and the cancer patient. Cancer Invest 2008 Mar;26(2):202-207 [FREE Full text] [doi: 10.1080/07357900701566197] [Medline: 18259953]
36. Lustria ML, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
37. Rainie L, Fox S, Horrigan J, Lenhart A, Spooner T. Pew Internet. Tracking online life: How women use the Internet to cultivate relationships with family and friends URL: https://www.pewinternet.org/2000/05/10/tracking-online-life-how-women-use-the-internet-to-cultivate-relationships-with-family-and-friends/ [accessed 2019-06-04] [WebCite Cache ID 78sjogrQG]
38. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 9;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]
39. Ajzen I, Albarracín D, Hornik R. Prediction and Change of Health Behavior: Applying the Reasoned Action Approach. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2007.

40. Dixon G, Clarke C. The effect of falsely balanced reporting of the autism-vaccine controversy on vaccine safety perceptions and behavioral intentions. Health Educ Res 2013 Apr;28(2):352-359. [doi: 10.1093/her/cys110] [Medline: 23193194]
41. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
42. Hinnant A, Jenkins J, Subramanian R. Health journalist role conceptions. Journalism Practice 2015 Aug 14;10(6):763-781 [FREE Full text] [doi: 10.1080/17512786.2015.1053509]
43. Gesser-Edelsburg A, Shir-Raz Y. Risk Communication and Infectious Diseases in an Age of Digital Media. London: Routledge; 2016.
44. Gesser-Edelsburg A, Walter N, Shir-Raz Y. The “new public” and the “good ol’press”: Evaluating online news sources during the 2013 polio outbreak in Israel". Health Commun 2017 Dec;32(2):169-179. [doi: 10.1080/10410236.2015.1110224] [Medline: 27191949]
45. Sudau F, Friede T, Grabowski J, Koschack J, Makedonski P, Himmel W. Sources of information and behavioral patterns in online health forums: observational study. J Med Internet Res 2014 Jan 14;16(1):e10 [FREE Full text] [doi: 10.2196/jmir.2875] [Medline: 24425598]
46. Nasrabadi AN, Sabzevari S, Bonabi TN. Iranian women's experiences of health information seeking barriers: a qualitative study in Kerman. Iran Red Crescent Med J 2015 Feb;17(2):e25156 [FREE Full text] [doi: 10.5812/ircmj.25156] [Medline: 25834743]
47. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
48. Oh HJ, Lee B. The effect of computer-mediated social support in online communities on patient empowerment and doctor-patient communication. Health Commun 2012;27(1):30-41. [doi: 10.1080/10410236.2011.567449] [Medline: 21797714]
49. Fox S, Duggan M. Pew Internet. Washington, DC: Pew Research Center; 2013. Health online URL: https://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-06-04] [WebCite Cache ID 78sl7It0a]
50. Tustin N. The role of patient satisfaction in online health information seeking. J Health Commun 2010 Jan;15(1):3-17. [doi: 10.1080/10810730903465491] [Medline: 20390974]
51. Vance K, Howe W, Dellavalle RP. Social internet sites as a source of public health information. Dermatol Clin 2009 Apr;27(2):133-6, vi. [doi: 10.1016/j.det.2008.11.010] [Medline: 19254656]
52. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
53. Domanski MC, Cavale N. Self-reported. Aesthetic Plast Surg 2012 Dec;36(6):1292-1295. [doi: 10.1007/s00266-012-9977-z] [Medline: 23052381]
54. Kelly B, Hornik R, Romantan A, Schwartz JS, Armstrong K, DeMichele A, et al. Cancer information scanning and seeking in the general population. J Health Commun 2010 Oct;15(7):734-753 [FREE Full text] [doi: 10.1080/10810730.2010.514029] [Medline: 21104503]
55. Li Y, Wang X, Lin X, Hajli M. Seeking and sharing health information on social media: a net valence model and cross-cultural comparison. Technol Forecast Soc Change 2018 Jan;126:28-40 [FREE Full text] [doi: 10.1016/j.techfore.2016.07.021]
56. Szychta P, Zieliński T, Rykała J, Witmanowski H, Kruk-Jeromin J. Role of the internet in communication between patient and surgeon before rhinoplasty. J Plast Surg Hand Surg 2012 Sep;46(3-4):248-251. [doi: 10.3109/2000656X.2012.697376] [Medline: 22909242]
57. Yang Q, Beatty M. A meta-analytic review of health information credibility: belief in physicians or belief in peers? Health Inf Manag 2016 Aug;45(2):80-89. [doi: 10.1177/1833358316639432] [Medline: 27105476]
58. Beaudoin CE, Hong T. Health information seeking, diet and physical activity: an empirical assessment by medium and critical demographics. Int J Med Inform 2011 Aug;80(8):586-595. [doi: 10.1016/j.ijmedinf.2011.04.003] [Medline: 21640643]
59. Lee C, Ramírez AS, Lewis N, Gray SW, Hornik RC. Looking beyond the internet: examining socioeconomic inequalities in cancer information seeking among cancer patients. Health Commun 2012;27(8):806-817 [FREE Full text] [doi: 10.1080/10410236.2011.647621] [Medline: 22356137]
60. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
61. Li Y, Wang X. Seeking health information on social media: a perspective of trust, self-determination, and social support. J Organ End User Comput 2018;30(1):1-22. [doi: 10.4018/JOEUC.2018010101]
62. Clarke MA, Moore JL, Steege LM, Koopman RJ, Belden JL, Canfield SM, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: a literature review. Health Informatics J 2016 Dec;22(4):992-1016. [doi: 10.1177/1460458215602939] [Medline: 26377952]
63. Hou J, Shim M. The role of provider-patient communication and trust in online sources in internet use for health-related activities. J Health Commun 2010;15(Suppl 3):186-199. [doi: 10.1080/10810730.2010.522691] [Medline: 21154093]

64. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
65. Wong DK, Cheung MK. Online health information seeking and eHealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
66. Lin WY, Zhang X, Song H, Omori K. Health information seeking in the Web 2.0 age: trust in social media, uncertainty reduction, and self-disclosure. Comput Hum Behav 2016 Mar;56:289-294 [FREE Full text] [doi: 10.1016/j.chb.2015.11.055]
67. Townsend A, Leese J, Adam P, McDonald M, Li LC, Kerr S, et al. eHealth, participatory medicine, and ethical care: a focus group study of patients' and health care providers' use of health-related internet information. J Med Internet Res 2015 Jun 22;17(6):e155 [FREE Full text] [doi: 10.2196/jmir.3792] [Medline: 26099267]
68. Chen X, Hay JL, Waters EA, Kiviniemi MT, Biddle C, Schofield E, et al. Health literacy and use and trust in health information. J Health Commun 2018;23(8):724-734. [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]
69. Chen X, Orom H, Hay JL, Waters EA, Schofield E, Li Y, et al. Differences in rural and urban health information access and use. J Rural Health 2018 Nov 16:1-13. [doi: 10.1111/jrh.12335] [Medline: 30444935]
70. Kelley MS, Su D, Britigan DH. Disparities in health information access: results of a county-wide survey and implications for health communication. Health Commun 2016;31(5):575-582. [doi: 10.1080/10410236.2014.979976] [Medline: 26452300]
71. Poínhos R, Oliveira BM, van der Lans IA, Fischer AR, Berezowska A, Rankin A, et al. Providing personalised nutrition: consumers' trust and preferenceing sources of information, service providers and regulators, and communication channelss regard. Public Health Genomics 2017;20(4):218-228. [doi: 10.1159/000481357] [Medline: 29050032]
72. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Dec 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
73. Stevenson FA, Kerr C, Murray E, Nazareth I. Information from the internet and the doctor-patient relationship: the patient perspective--a qualitative study. BMC Fam Pract 2007 Aug 16;8:47 [FREE Full text] [doi: 10.1186/1471-2296-8-47] [Medline: 17705836]
74. Mendes Á, Abreu L, Vilar-Correia MR, Borlido-Santos J. "That Should Be Left to Doctors, That's What They are There For!"-Exploring the Reflexivity and Trust of Young Adults When Seeking Health Information. Health Commun 2017 Dec;32(9):1076-1081. [doi: 10.1080/10410236.2016.1199081] [Medline: 27485762]
75. Metzger MJ, Flanagin AJ. Using Web 2.0 technologies to enhance evidence-based medical information. J Health Commun 2011;16(Suppl 1):45-58. [doi: 10.1080/10810730.2011.589881] [Medline: 21843095]
76. Yang ZJ, Kahlor LA. What, me worry? the role of affect in information seeking and avoidance. Science Communication 2012 May 24;35(2):189-212 [FREE Full text] [doi: 10.1177/1075547012441873]
77. Ye Y. Correlates of consumer trust in online health information: findings from the health information national trends survey. J Health Commun 2011 Jan;16(1):34-49. [doi: 10.1080/10810730.2010.529491] [Medline: 21086209]
78. Yun EK, Park H. Consumers' disease information-seeking behaviour on the internet in Korea. J Clin Nurs 2010 Oct;19(19-20):2860-2868. [doi: 10.1111/j.1365-2702.2009.03187.x] [Medline: 20598000]
79. Agree EM, King AC, Castro CM, Wiley A, Borzekowski DL. "It's got to be on this page": age and cognitive style in a study of online health information seeking. J Med Internet Res 2015 Mar 24;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]
80. Cherla DV, Sanghvi S, Choudhry OJ, Jyung RW, Eloy JA, Liu JK. Readability assessment of Internet-based patient education materials related to acoustic neuromas. Otol Neurotol 2013 Sep;34(7):1349-1354. [doi: 10.1097/MAO.0b013e31829530e5] [Medline: 23921935]
81. Finnie RK, Felder TM, Linder SK, Mullen PD. Beyond reading level: a systematic review of the suitability of cancer education print and Web-based materials. J Cancer Educ 2010 Dec;25(4):497-505 [FREE Full text] [doi: 10.1007/s13187-010-0075-0] [Medline: 20237884]
82. Misra P, Kasabwala K, Agarwal N, Eloy JA, Liu JK. Readability analysis of internet-based patient information regarding skull base tumors. J Neurooncol 2012 Sep;109(3):573-580. [doi: 10.1007/s11060-012-0930-4] [Medline: 22810759]
83. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the internet and e-mail for health care information: results from a national survey. J Am Med Assoc 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
84. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
85. Dolan G, Iredale R, Williams R, Ameen J. Consumer use of the internet for health information: a survey of primary care patients. Int J Cons Stud 2004 Mar;28(2):147-153 [FREE Full text] [doi: 10.1111/j.1470-6431.2003.00363.x]
86. Dutta-Bergman MJ. Primary sources of health information: comparisons in the domain of health attitudes, health cognitions, and health behaviors. Health Commun 2004;16(3):273-288. [doi: 10.1207/S15327027HC1603_1] [Medline: 15265751]
87. Journalism.org. Washington, DC: Pew Research Center; 2010. How news happens: a study of the news ecosystem of one American city URL: https://www.journalism.org/2010/01/11/how-news-happens/ [accessed 2019-06-04] [WebCite Cache ID 78soDJ0Fz]

88. Geana MV, Kimminau KS, Greiner KA. Sources of health information in a multiethnic, underserved, urban community: does ethnicity matter? J Health Commun 2011 Jul;16(6):583-594. [doi: 10.1080/10810730.2011.551992] [Medline: 21391043]
89. Tu HT. Surprising decline in consumers seeking health information. Track Rep 2011 Nov(26):1-6. [Medline: 22121566]
90. Avery EJ, Lariscy RW, Sohn Y. Public information officers' and journalists' perceived barriers to providing quality health information. Health Commun 2009 Jun;24(4):327-336. [doi: 10.1080/10410230902889365] [Medline: 19499426]
91. Gaglio B, Glasgow RE, Bull SS. Do patient preferences for health information vary by health literacy or numeracy? A qualitative assessment. J Health Commun 2012;17(Suppl 3):109-121. [doi: 10.1080/10810730.2012.712616] [Medline: 23030565]
92. Brezis M. Who will look after the health of the media? Zman Harefuah 2002;1(6):44-46 [FREE Full text]
93. Bubela TM, Caulfield TA. Do the print media. CMAJ 2004 Apr 27;170(9):1399-1407 [FREE Full text] [Medline: 15111473]
94. Høye S, Hjortdahl P. [. Tidsskr Nor Laegeforen 2002 Jun 30;122(17):1671-1676 [FREE Full text] [Medline: 12555610]
95. Moynihan R, Sweet M. Medicine, the media and monetary interests: the need for transparency and professionalism. Med J Aust 2000;173(11-12):631-634. [Medline: 11379512]
96. Ransohoff DF, Ransohoff RM. Sensationalism in the media: when scientists and journalists may be complicit collaborators. Eff Clin Pract 2001;4(4):185-188. [Medline: 11525108]
97. Shuchman M, Wilkes MS. Medical scientists and health news reporting: a case of miscommunication. Ann Intern Med 1997 Jun 15;126(12):976-982. [Medline: 9182476]
98. Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. N Engl J Med 2000 Jun 1;342(22):1645-1650. [doi: 10.1056/NEJM200006013422206] [Medline: 10833211]
99. Tanner AH, Friedman DB, Zheng Y. Influences on the construction of health news: the reporting practices of local television news health journalists. J Broadcast Electron Media 2015 May 28;59(2):359-376 [FREE Full text] [doi: 10.1080/08838151.2015.1029123]
100. Weitkamp E. On the roles of scientists, press officers and journalists. J Sci Commun 2014;13(03):1-3 [FREE Full text] [doi: 10.22323/2.13030501]
101. Foster C, Tanner AH, Kim SH, Kim SY. National conversations about the costs of US health care. Sci Commun 2014 Jun 22;36(5):519-543. [doi: 10.1177/1075547014536882]
102. Shuyler KS, Knight KM. What are patients seeking when they turn to the internet? Qualitative content analysis of questions asked by visitors to an orthopaedics web site. J Med Internet Res 2003 Oct 10;5(4):e24 [FREE Full text] [doi: 10.2196/jmir.5.4.e24] [Medline: 14713652]
103. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. J Am Med Assoc 2002;287(20):2691-2700. [Medline: 12020305]
104. Meric F, Bernstam EV, Mirza NQ, Hunt KK, Ames FC, Ross MI, et al. Breast cancer on the world wide web: cross sectional survey of quality of information and popularity of websites. Br Med J 2002 Mar 9;324(7337):577-581 [FREE Full text] [doi: 10.1136/bmj.324.7337.577] [Medline: 11884322]
105. Boyer C, Gaudinat A, Baujard V, Geissbühler A. Health on the Net Foundation: assessing the quality of health web pages all over the world. Stud Health Technol Inform 2007;129(Pt 2):1017-1021. [Medline: 17911869]
106. 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 Jun 28;314(7098):1875-1879 [FREE Full text] [doi: 10.1136/bmj.314.7098.1875] [Medline: 9224132]
107. Provost M, Koompalum D, Dong D, Martin BC. The initial development of the WebMedQual scale: domain assessment of the construct of quality of health web sites. Int J Med Inform 2006 Jan;75(1):42-57. [doi: 10.1016/j.ijmedinf.2005.07.034] [Medline: 16169770]
108. Godin S, Truschel J, Singh V. Quality assurance of self-help sites on the internet. J Prev Interv Community 2005 Jan 25;29(1-2):67-84 [FREE Full text] [doi: 10.1300/J005v29n01_05]
109. Dalton DM, Kelly EG, Molony DC. Availability of accessible and high-quality information on the internet for patients regarding the diagnosis and management of rotator cuff tears. J Shoulder Elbow Surg 2015 May;24(5):e135-e140. [doi: 10.1016/j.jse.2014.09.036] [Medline: 25457189]
110. Leira-Feijoo Y, Ledesma-Ludi Y, Seoane-Romero JM, Blanco-Carrión J, Seoane J, Varela-Centelles P. Available web-based dental implants information for patients. How good is it? Clin Oral Implants Res 2015 Nov;26(11):1276-1280. [doi: 10.1111/clr.12451] [Medline: 25041413]
111. Mozafarpour S, Norris B, Borin J, Eisner BH. Assessment of readability, quality and popularity of online information on ureteral stents. World J Urol 2018 Jun;36(6):985-992. [doi: 10.1007/s00345-018-2179-9] [Medline: 29435639]
112. Ritchie L, Tornari C, Patel PM, Lakhani R. Glue ear: how good is the information on the World Wide Web? J Laryngol Otol 2016 Feb;130(2):157-161. [doi: 10.1017/S0022215115003230] [Medline: 26805389]
113. Sacchi M, Yeung TM, Spinelli A, Mortensen NJ. Assessment of the quality of patient-orientated internet information on surgery for ulcerative colitis. Colorectal Dis 2015 Jun;17(6):511-514. [doi: 10.1111/codi.12870] [Medline: 25510335]

114. Caiata-Zufferey M, Abraham A, Sommerhalder K, Schulz PJ. Online health information seeking in the context of the medical consultation in Switzerland. Qual Health Res 2010 Aug;20(8):1050-1061. [doi: 10.1177/1049732310368404] [Medline: 20442347]
115. Hu X, Bell RA, Kravitz RL, Orrange S. The prepared patient: information seeking of online support group members before their medical appointments. J Health Commun 2012;17(8):960-978. [doi: 10.1080/10810730.2011.650828] [Medline: 22574697]
116. Vehovar V, Toepoel V, Steinmetz S. Non-probability sampling. In: Wolf C, Joye D, Smith TW, Fu YC, editors. The SAGE Handbook of Survey Methodology. London: Sage Publications; 2016:329-345.
117. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Methods Meas 2007 Apr;1(1):77-89 [FREE Full text] [doi: 10.1080/19312450709336664]
118. Ulin P, Robinson ET, Tolley EE. Qualitative Methods in Public Health: A Field Guide for Applied Research. San Francisco, CA: Jossey-Bass; 2005.
119. Kreps GL. The pervasive role of information in health and health care: implications for health communication policy. Ann Int Commun Assoc 2016 May 18;11(1):238-276 [FREE Full text] [doi: 10.1080/23808985.1988.11678690]
120. Rubin A, Rubin B. Interface of personal and mediated communication: fifteen years later. Electr J Commun 2001;11(1):1 [FREE Full text]
121. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, et al. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. Health Info Libr J 2013 Sep;30(3):178-190 [FREE Full text] [doi: 10.1111/hir.12036] [Medline: 23981019]
122. Cao YG, Cimino JJ, Ely J, Yu H. Automatically extracting information needs from complex clinical questions. J Biomed Inform 2010 Dec;43(6):962-971 [FREE Full text] [doi: 10.1016/j.jbi.2010.07.007] [Medline: 20670693]
123. Norbert GL, Lwoga ET. Information seeking behaviour of physicians in Tanzania. Inf Dev 2012 Jul 9;29(2):172-182 [FREE Full text] [doi: 10.1177/0266666912450449]
124. Randell R, Mitchell N, Thompson C, McCaughan D, Dowding D. From pull to push: understanding nurses' information needs. Health Informatics J 2009 Jun;15(2):75-85. [doi: 10.1177/1460458209102969] [Medline: 19474221]
125. Mårtensson L, Hensing G. Health literacy - a heterogeneous phenomenon: a literature review. Scand J Caring Sci 2012 Mar;26(1):151-160 [FREE Full text] [doi: 10.1111/j.1471-6712.2011.00900.x] [Medline: 21627673]
126. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012 Sep 24;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
127. Wicker A. Attitudes versus actions: the relationship of verbal and overt behavioral responses to attitude objects. J Soc Issues 1969;25(4):41-78 [FREE Full text] [doi: 10.1111/j.1540-4560.1969.tb00619.x]
128. McNab C. What social media offers to health professionals and citizens. Bull World Health Organ 2009 Aug;87(8):566 [FREE Full text] [Medline: 19704998]
129. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. Health Promot Pract 2013 Mar;14(2):157-162. [doi: 10.1177/1524839912469378] [Medline: 23271716]
130. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. Health Commun 2012;27(2):179-185. [doi: 10.1080/10410236.2011.571759] [Medline: 21827326]
131. Thackeray R, Neiger BL, Burton SH, Thackeray CR. Analysis of the purpose of state health departments' tweets: information sharing, engagement, and action. J Med Internet Res 2013 Nov 11;15(11):e255 [FREE Full text] [doi: 10.2196/jmir.3002] [Medline: 24217361]
132. Thackeray R, Neiger BL, Smith AK, van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012 Mar 26;12(1):242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
133. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981 Jan 30;211(4481):453-458. [Medline: 7455683]
134. Gesser-Edelsburg A, Walter N, Green MS. Health care workers-part of the system or part of the public? Ambivalent risk perception in health care workers. Am J Infect Control 2014 Aug;42(8):829-833. [doi: 10.1016/j.ajic.2014.04.012] [Medline: 24939516]
135. Heimberger T, Chang HG, Shaikh M, Crotty L, Morse D, Birkhead G. Knowledge and attitudes of healthcare workers about influenza: why are they not getting vaccinated? Infect Control Hosp Epidemiol 1995 Jul;16(7):412-415. [Medline: 7673647]
136. Marshall RJ. Influenza vaccine use among health care workers. Soc Market Q 2013 Sep 24;19(4):222-229 [FREE Full text] [doi: 10.1177/1524500413505096]
137. Smedley J, Poole J, Waclawski E, Stevens A, Harrison J, Watson J, et al. Influenza immunisation: attitudes and beliefs of UK healthcare workers. Occup Environ Med 2007 Apr;64(4):223-227 [FREE Full text] [doi: 10.1136/oem.2005.023564] [Medline: 17182640]
138. Corcoran N. Communicating Health: Strategies for Health Promotion. Thousand Oaks, CA: Sage Publications; 2013.

139. Du Pré A. Communicating About Health: Current Issues and Perspectives. Mountain View, CA: Mayfield Publishing Company; 2000.
140. Morgan MG, Fischhoff B, Bostrom A, Atman CJ. Risk Communication: A Mental Models Approach. Cambridge: Cambridge University Press; 2002.
141. Winsten JA. Science and the media: the boundaries of truth. Health Aff (Millwood) 1985;4(1):5-23. [Medline: 3997047]
142. Caulfield T. The commercialisation of medical and scientific reporting. PLoS Med 2004 Dec;1(3):e38 [FREE Full text] [doi: 10.1371/journal.pmed.0010038] [Medline: 15630461]
143. Kua E, Reder M, Grossel M. Science in the news: a study of reporting genomics. Public Underst Sci 2016 Aug 18;13(3):309-322 [FREE Full text] [doi: 10.1177/0963662504045539]
144. Cook DM, Boyd EA, Grossmann C, Bero LA. Reporting science and conflicts of interest in the lay press. PLoS One 2007 Dec 05;2(12):e1266 [FREE Full text] [doi: 10.1371/journal.pone.0001266] [Medline: 18060060]
145. Schwitzer G. How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. PLoS Med 2008 May 27;5(5):e95 [FREE Full text] [doi: 10.1371/journal.pmed.0050095] [Medline: 18507496]


## Abbreviations

**HCW:** health care worker


*Edited by G Eysenbach; submitted 22.03.19; peer-reviewed by S Paige, X Chen, M Torii; comments to author 17.04.19; revised version received 24.05.19; accepted 29.05.19; published 03.07.19.*

*Please cite as:*
*Gesser-Edelsburg A, Abed Elhadi Shahbari N, Cohen R, Mir Halavi A, Hijazi R, Paz-Yaakobovitch G, Birman Y*
*Differences in Perceptions of Health Information Between the Public and Health Care Professionals: Nonprobability Sampling Questionnaire Survey*
*J Med Internet Res 2019;21(7):e14105*
*URL: https://www.jmir.org/2019/7/e14105/*
*doi:10.2196/14105*
*PMID:31271145*

Original Paper

# Users' Experiences With Web-Based Health Care Information: Qualitative Study About Diabetes and Dementia Information Presented on a Governmental Website

Therese Agnes Wiegers[1], PhD; Michelle Hendriks[2], PhD; Uriëll Malanda[1], PhD; Dolf de Boer[1], PhD

[1]Nivel, Utrecht, Netherlands

[2]Reinier van Arkel, 's Hertogenbosch, Netherlands

**Corresponding Author:**
Dolf de Boer, PhD
Nivel
Otterstraat 118-124
3513CR
Utrecht,
Netherlands
Phone: 33 30 2729 658
Email: D.deBoer@nivel.nl

## Abstract

**Background:** Information on health and health care is abundant on the internet. To make informed choices, patients need reliable and easy-to-understand information about quality and availability of care providers and treatment options. However, the reliability of such Web-based information is difficult to assess.

**Objective:** This study aimed to test Web-based information about diabetes and dementia and specifically a new presentation format of care routes to see if people are able to understand and use the information.

**Methods:** Overall, 38 cognitive interviews were held; 20 people viewed the information about diabetes and 18 viewed the dementia information. Participants were asked what they would want to know about either diabetes or dementia, what choices they would want to make concerning their preferred care provider and treatment, and what information they would like to find to make these choices. They were then asked to view the relevant pages and comment on them. The interview was focused on general information about the condition, the care route, and the quality information for choosing a hospital. The interviews were transcribed verbatim and then systematically coded and ordered into themes.

**Results:** The themes that were developed for both Web pages during the analysis were information needs, findability, usability, comprehension and readability, recognizability, care route, quality information, and usefulness. Information needs were found to be very diverse and dependent on the personal situation and condition of the participant. Several participants were unable to find specific items because they were not where they expected them to be. Most participants were positive about the layout, font, and color scheme of the test pages. However, options of clicking through to another website and indications where information can be expanded and collapsed could be made clearer. Participants generally found the information easy to understand but felt a need for a more detailed explanation of the medical terms. Recognition of the information played an important role: participants assessed whether the information they found matched their experiences. The term care route meant little to most of the participants, but the layout of the care route itself was found to be clear. Not many respondents spontaneously went to the quality information, and a number of participants had difficulty understanding it. Overall, the participants thought the information on the website was useful and helpful.

**Conclusions:** The cognitive interviews gave numerous insights into how Web-based information is processed and understood. The care route offers a clear overview of the various stages as the condition progresses, but the name *care route* is not clear to everyone. We gained insight into differences between subgroups of people in terms of information needs, comprehension, and use of the information because the diversity within the group of participants was lower than expected.

***(J Med Internet Res 2019;21(7):e11340)*** doi:10.2196/11340

**KEYWORDS**

patients; qualitative research; choice behavior

# Introduction

## Background

Information on health and health care is abundant on the internet; it is present in all kinds of forms. Health information can be seen as information about diseases or health conditions; about signs and symptoms; about progression, severity, and duration; and about treatment possibilities. Health care information is more about care providers and care institutions, about accessibility and availability, about costs and quality, and about choosing a care provider or a treatment.

It is well understood that to be able to make an informed choice for a specific care provider or a specific treatment, patients need reliable and easy-to-understand information about the quality and availability of care providers and treatment options. However, the reliability of much of the information found on the internet is difficult to assess. In addition, the availability of information alone is no guaranty for its proper use. For instance, free provider choice and costs are relevant as well [1]. Governments or providers or insurers may be concerned about the reliability of the information available on the internet and may want to provide some kind of guidance to consumers looking for information. This may have 2 purposes: informing consumers about specific health issues by providing reliable and up-to-date medical information (providing health information) or informing consumers about quality issues to encourage them to make informed choices between care providers, care institutions, or treatment options (providing health care information or choice information). Websites may focus on one of these aims or may be designed for both uses. NHS Choices [2] is an example of a government website providing health information as well as choice information. A recent study by Lee et al has shown that there is a need among health information seekers for a centralized health portal or database containing relevant and quality health information [3].

In the previous 20-odd years, several studies have been conducted about the way information about health and health care could best be presented on the internet to make it accessible for anyone. According to Eysenbach [4], there are different levels of accessibility. The first level is physical accessibility: having access to the information, in this case, having a computer and having access to the internet. The second level is findability: finding the relevant website or finding information within a website. The third level is readability: is the language understandable for most people. The final level of accessibility is usability: the way the information is grouped and presented, how the user navigates through the information, and the amount of help the system gives.

Some studies focused on the information-seeking behavior of health consumers [5] or of health consumers and health professionals [6]. Pang et al discussed different search approaches in the process of health information seeking, depending on the level of knowledge about the health problem, level of curiosity, and perceived situational relevance to the health problem [7]. They state that people familiar with the health problem will show different seeking behavior than people not familiar with the condition. A review by Victoor et al [8] was focused on whether and how patients choose a provider. Other studies specifically focused on providing comparative health care information to consumers, with the intention to facilitate choice, such as American studies concerning decisions about which health plan to join [9,10,11]. As many of these studies have shown [12,13], the way choice information is presented is of utmost importance. Years earlier, Bettman and Kakkar [14] already examined the effect of information presentation format on consumers' information acquisition strategies. They found that the strategies used to acquire information are strongly affected by the structure of the information presented. These results have implications for decisions on how to present information to consumers [14]. Their findings also apply to health care information.

Providing consumers with reliable medical information about specific health issues requires a different approach than providing choice information. Conflicting views and standards exist on how (and how much) medical information should be provided to consumers [4]. What is known is that people on average can process about 6 pieces of information at a time and are easily overwhelmed by information [15]. Research on how information is presented invariably shows that information on a website should be kept as simple as possible [12,13,16]. Information should be provided in layers, with a clear but concise overview on the main page, with links to additional information for those who are interested [16]. However, even with the information available, only a small number of people will use that to make an informed choice [1,8].

The Dutch government sees it as its responsibility to make understandable and reliable information about the quality and availability of health care in the Netherlands accessible for everyone. In 2004, the Dutch Ministry of Health commissioned the National Institute for Public Health and the Environment to develop, host, and manage a public national health and care portal on the internet, called KiesBeter, which translates as *Choose Better*. Its objectives were twofold: to improve the information position of consumers and to enable citizens to make informed decisions in health, health care, and health insurance. The portal seeks to provide reliable, independent, and coherent information on a range of health care–related topics [17]. Since 2014, the portal is hosted by the National Healthcare Institute of the Netherlands (Zorginstituut Nederland) [18], which is tasked with making information about the quality of care available to the public. The idea is that this portal will function as a starting point for health care consumers and will direct people through links to reliable and relevant information, provided by other sources, such as health care professionals, patient organizations, or health insurance companies. On the website, KiesBeter, people can find general information about health conditions, guidelines for good quality health care in general and for the appropriate care for any given condition, and comparative information to choose a health care provider or hospital [19].

Since the update in 2014, the website has experimented with providing information in the form of *care routes* for specific conditions. *Care routes* are not the same as *care pathways*. *Care pathways* are usually presented as a description of the necessary care for an average patient with a certain condition and are mostly aimed at health care providers [20]. The *care route*, as presented on KiesBeter, shows the trajectory of a typical patient through the health care system, from diagnosis to possible end stages as the illness progresses, with appropriate links to other websites with reliable information about the condition and treatment in that particular stage. Visually, a *care route* is presented as a kind of railway track, with stations representing access to specific information or available choices (see Multimedia Appendix 1). Thus, the *care route* combines health information and health care information. The aim of presenting *care routes* was to provide visitors of the KiesBeter website with insights into what good care is for a particular condition, what choices of treatment or of care providers they will encounter, and when they will encounter them and to encourage them to make their own choices based on high-quality information available on other websites.

Presenting *care routes* on the internet is a new phenomenon in Dutch health care and elsewhere. Little is known about the effectiveness and usability for health care consumers of information presented this way. In 2013, the Netherlands Institute for Health Services Research (NIVEL) conducted a study to find evidence in (international) literature that a *care route* is a useful format for encouraging people to choose between care providers [16]. The conclusion was that the proposed presentation of a *care route* on KiesBeter fits nicely with the requirements (as simple as possible, in layers, and with links to additional information) and may encourage patients to make informed choices about care providers [16].

## Objectives

Before actually implementing the *care route* for a number of conditions, an additional study was conducted to test the format among patients to see if they understand the information and are able to use it. In this paper, we present the results of that study carried out in 2015. For this study, information pages including *care routes* for dementia and diabetes were made available as test versions that were not yet available on the dementia and diabetes pages on the KiesBeter website. Cognitive interviews were held with a range of consumers, aiming to get a clear picture of users' experiences with these pages on KiesBeter, with a focus on what information is sought (information needs) and how the information is understood (information comprehension).

The research answers the following questions:

1. What information do the majority of participants want to find and what information is only read by a few participants? (Information needs.)
2. How do participants read and process the information on the KiesBeter website, in particular, the *care routes* for dementia and diabetes? (Information comprehension.)

# Methods

## Participants

Participants were recruited through various channels with the aim of reaching different groups in society:

- NIVEL invited 246 members of the Dutch Health Care Consumer Panel to participate in a personal interview. The selection criteria were location (not too far from the place where the interviews were held), patient activation score, health literacy score, and education level. Members with a low patient activation score were oversampled. Familiarity with diabetes or dementia was not included in the selection criteria. This resulted in 11 interviews, 10 focusing on the diabetes page and 1 on the dementia page. Most of them were familiar with diabetes (6) or dementia (1), whereas this is unknown for the remaining 4.
- The Dutch Diabetes Association (DVN) invited people to participate via their website, their newsletter, and Twitter. This resulted in 10 interviews, all focusing on the diabetes page.
- Alzheimer Nederland sent 193 members of the Dutch Alzheimer Panel invitations by email for an interview. This panel consists of informal carers of people with dementia. This resulted in 17 interviews focusing on the dementia page.

We purposefully recruited participants from DVN and Alzheimer Nederland because they are familiar with the condition, so they would be motivated to look for information that would be important or relevant to them.

## Structure of the Diabetes and Dementia Web Pages

Each Web page opened with general information about the condition ordered in short statements or questions that could be expanded. One of those statements was *care route*. When expanding that statement, the *care route* became visible, divided in *stations*, which, in turn, could be expanded to reveal more information and link to a different website (see Multimedia Appendix 1). For diabetes, the main *stations* were diagnosis, first treatment phase, and chronic treatment phase. The route between the main stations showed additional stations. For instance, the stations after diagnosis were complaints, early detection, and elevated blood sugar value, each of which could be expanded for additional information. Following the statements was a text section called "Find a care provider," which linked to another part of the website, where on the basis of distance (from a given position) and a choice of quality indicators (not obligatory), up to 3 hospitals could be selected. On the right-hand side of the page, in a different color, additional information was presented about the quality of care for the relevant condition and about the prevalence of the condition. On the diabetes page, the "Find a care provider" link was repeated here.

## Structure of the Cognitive Interview

In the interviews, before opening the Web page, the participants were first asked what they would want to know about either diabetes or dementia, what choices they would want to make concerning their preferred care provider and possible treatment, and what information they would like to find to make these choices. They were then asked to view the relevant KiesBeter pages and see if they could find the answers to those questions.

The interview consisted of 3 parts: part 1 was focused on the general information about diabetes or dementia, part 2 was focused on the *care route*, and part 3 was focused on the quality information for choosing a hospital. We used the thinking-aloud technique combined with verbal probing. The participants were asked to view the Web page as if they were at home and think out loud while doing so. They were invited to describe what they saw and to explain certain terms used and infographics presented on the website. Verbal probing was used to clarify how the information was being understood and interpreted and what might be missed. Each part of the interview and the interview as a whole was concluded by asking the participants if they had been able to find the information they were looking for, if any information was missing on the website, and if they had any suggestions for improvements. After each interview, the interviewers wrote down their impression of the conversation and any significant observations.

### Analyses

The interviews were transcribed verbatim and then systematically coded and ordered into themes by 2 researchers (MH and a junior researcher). The researchers conducted descriptive thematic analysis, consisting of an open coding and an axial coding phase: relevant themes were extracted, categorized, and classified. Then, relationships between categories were identified in the axial coding phase [21]. The results below are ordered according to these themes. Where relevant, a distinction is made by condition: diabetes or dementia.

## Results

### Participants

A total of 38 people participated in the cognitive interviews: 20 people, of whom 14 were diagnosed with diabetes or were familiar with the condition, viewed the page about diabetes and 18 people, all familiar with dementia, viewed the dementia page. The interviews were held at NIVEL or in people's homes. Table 1 gives an overview of the participants. The average age was 65.7 years (ranging from 29 to 81; SD 10.7); 65% of them were older than 65 years. A little over half of them were women. The participants were highly educated compared with the general population; almost half of them had completed higher professional education or university education. One-third (32%) of the participants said their health was moderate or poor. Three-quarters of the participants had been using the internet for over 8 years, and 3 out of 4 of them used it daily. Less than half of the participants had visited the KiesBeter website before participating in the study. The majority of the participants considered themselves to be good with computers, felt confident in their use of the internet for collecting information, and said they used the internet to find information that was difficult to find elsewhere.

Our invitation policy had been aimed at different groups regarding the level of education, health literacy, patient activation, internet use, and health status. However, as Table 1 shows, the people responding turned out to be a selected group of, on average, highly educated, healthy, and confident internet users.

### General Impression

A finding from the interviewers' observations was that it was difficult for the respondents to view all parts of the selected Web page within the time frame of 1 hour. This means that not all the elements of the Web page have always been addressed to their full depth. The website contained too much information to view and discuss in the allotted time. In addition, it was noted that the participants varied considerably in the extent in which they thought aloud when viewing the website. Some talked spontaneously about what they saw and thought of the website, whereas others needed more prompting. Their internet skills also varied, as did the way the participants browsed the website. Where some might systematically view the Web page from top to bottom and read all the available information, others would scroll quickly through the information and only read what they were interested in. Some people only looked at the information on KiesBeter, others also clicked through to other websites. Therefore, the general impression was that although the participants were a selected group regarding their education, health, and internet use, they were a very diverse group with regard to their behavior while viewing and commenting on the Web pages under study.

### Themes

The themes for both Web pages that were developed during the descriptive thematic analysis were: information needs, findability, usability, understanding and readability, recognizability, *care route*, quality information, and usefulness. Some of these themes (findability, usability, understanding and readability, and recognizability) closely resemble the earlier levels of accessibility of information described by Eysenbach [3].

#### *Information Needs*

The information needs were very diverse and dependent on the personal situation and condition of the participant. Some of the participants did not have a clear question at the time of the interview because they had been coping with the condition for years and already were familiar with most of the information. Several participants also mentioned that they already received all the information they wanted from their health care providers, such as their general practitioner, case manager, or specialist. Informal caregivers of people with dementia also regularly mentioned the Alzheimer Café and the Alzheimer Nederland website as sources of information:

> *Yes, actually I already know a lot about diabetes, so yeah. No, I keep up with new information, I pretty much already know the rest.*
>
> *"What is dementia"...I'd skip that bit, because I think I already know.*

**Table 1.** Characteristics of the cognitive interview participants for KiesBeter (N=38).

| Demographic characteristic | Statistics, n (%) |
|---|---|
| **Gender** | |
| Men | 17 (45) |
| Women | 21 (55) |
| **Diabetes, dementia, or unknown** | |
| Diabetes patient or familiar with diabetes | 16 (42) |
| Familiar with dementia | 18 (47) |
| Unknown | 4 (11) |
| **Education** | |
| Lower or prevocational secondary education | 2 (5) |
| General secondary education | 5 (13) |
| Senior secondary vocational education | 7 (18) |
| Senior general secondary education and preuniversity education | 7 (18) |
| Higher vocational education | 9 (24) |
| University education | 8 (21) |
| **General health** | |
| Excellent | 1 (3) |
| Very good | 9 (24) |
| Good | 16 (42) |
| Moderate | 10 (26) |
| Poor | 2 (5) |
| **Years using the internet** | |
| Less than 4 | 3 (8) |
| 4-6 | 2 (5) |
| 6-8 | 4 (11) |
| More than 8 | 29 (76) |
| **Frequency of internet use** | |
| Daily | 28 (74) |
| Several times a week | 8 (21) |
| Once a week | 1 (3) |
| Once a month | 1 (3) |
| **Visited the KiesBeter website before** | |
| No | 21 (55) |
| Yes | 17 (45) |
| **I'm good with computers** | |
| (Entirely) Disagree | 4 (11) |
| Neither agree nor disagree | 10 (26) |
| (Entirely) Agree | 24 (63) |
| **I often surf the internet without really knowing what I'm looking for** | |
| (Entirely) Disagree | 24 (63) |
| Neither agree nor disagree | 8 (21) |
| (Entirely) Agree | 6 (16) |
| **I feel confident about using the internet to collect information** | |

| Demographic characteristic | Statistics, n (%) |
|---|---|
| (Entirely) Disagree | 0 (0) |
| Neither agree nor disagree | 7 (19) |
| (Entirely) Agree | 29 (81) |
| **Using the internet lets me find information that would be difficult to find elsewhere** | |
| (Entirely) Disagree | 2 (5) |
| Neither agree nor disagree | 9 (24) |
| (Entirely) Agree | 26 (70) |

### Diabetes

Participants who viewed the Web page about diabetes said they wanted information about the cause of diabetes, the different types of diabetes, the early symptoms and the progress of the disease, what they can do themselves to stay in control of the condition, and the latest developments in treatment options. A number of participants said that it should be made clear and stated in several places that the information on the website is only about type 2 diabetes and not type 1 diabetes:

> *The symptoms of early diabetes are actually non-existent or very limited. I'd add that.*
>
> *I'd like to know how you get diabetes and whether it is linked to genetics or it is something you might get because of your eating habits.*
>
> *I would look at what the current treatment methods are and...do I need syringes right away or do I get pills? That kind of thing.*
>
> *What I can do myself. How do I get a grip on diabetes. The healthy lifestyle.*
>
> *Well, more about new developments in diabetes care or whether there is light in the darkness. Even though I'm sixty-nine.*

### Dementia

The informal carers of people with dementia were also interested in information about the causes of the condition, its various forms, and its progress. Other aspects they mentioned were how to deal with someone with dementia, where they can get support for caring for their relative, the problems they may encounter, and various types of living arrangements. The informal carers also wanted practical information about available medical devices (such as smart home devices), where certain types of assistance can be requested (such as contact information for the care institutions), and how much this costs:

> *I would actually like to know if it's hereditary, because most of my dad's side of the family have had dementia.*
>
> *What types of dementia there are and...what consequences they have.*
>
> *It tells you here what you can do yourself, but I think there should be a referral to an informal-care support centre near you.*
>
> *What kinds of living arrangements are there? How do you get information about them?*

### *Findability*

The majority of participants said that they did not know about the KiesBeter website until they were invited to participate in this study. They also said that they probably would not have found it by searching the internet:

> *How do you find a website like this? I'd never type in "kies beter" [choose better] to find a website like this.*
>
> *I still think that I'd advertise it a bit more. It could be advertised on TV or a flyer—if you want information, just have a look at this website.*

With regard to being able to find information on the website, several participants were unable to find specific items because they were not where they expected them to be:

> *The diagnosis, yes...What symptoms do you have? The symptoms aren't mentioned here; I would have expected that though.*
>
> *This is what I expected to find in the 'care route'. Yeah, it's there, like I thought...but, it's too far away and hidden.*
>
> *Getting crisis assistance involved...But I wonder: shouldn't that also be part of 'living at home', because that is when you might need this information.*

### *Usability*

Most participants were positive about the layout, font, and color scheme of the website. Some said they would like a bigger font that stayed big when they clicked through to other pages or subpages:

> *The layout's fine. It's not too complicated. It's easy to read.*
>
> *Yes, it's nicely laid out. The colours are clear.*

People appreciated the fact that the website used short texts and then referred elsewhere for more information. However, it could be made more obvious where there is the option of clicking through to another website and where information on this website can be expanded and collapsed. It is important that this is clear and uniform throughout the entire website:

> *I like this, these links too. I like it because you keep an overview and can go back.*
>
> *I can get enough information now that you've shown me these click-through options, but I hadn't noticed them myself.*

> *I can see that those words are a lighter blue, but it'd be better if they were underlined or a different colour.*

When clicking on a link to another website (eg, to DVN or Alzheimer Nederland), the participants noted that they did not arrive on pages with the specific information they were looking for. Often, the link was to a home page, and they again had to search for information on that website. There was also a need for more links to other websites or subpages.

At the right-hand side of the Web page, a link was available, leading to quality information under the heading "Find a care provider." However, several participants did not see the "Find a care provider" block, mainly because people did not look at the right-hand side of the Web page. Some informal carers looked for the quality information in the *care route*, in the parts about the diagnosis, or in "When living at home becomes impossible."

The layout of the *care route* itself was felt to be clear. However, it was not clear to everyone that the bullets could be expanded (ie, when clicking a bullet, that part of the route was unfolded to present more detailed information):

> *I like this bit. You can expand this, yeah. That's nice and clear. And the text is legible, enough space between, and so forth.*

> *Sometimes there's a bit too much information on a single page.*

> *That's a bit cumbersome...That you can click the bullets. You could add 'hide' or 'show' next to them.*

### Understanding and Readability

The information was generally found to be easy to understand, but the participants felt a need for a more detailed explanation of the medical terms. This could be done by adding a reference (link) or an information button. The quality aspects in the quality information section could also be explained in more detail:

> *You can clearly read what you can expect to happen. That's something you should be reading, at any rate.*

> *When they use terms I'd like them to be explained.*

> *These words are so complicated, all those medical terms.*

Some participants felt the language was too distant and too technical:

> *The entire tone of the text…it turns me right off, to be honest. I don't like it at all, I feel like I'm being treated a bit condescendingly.*

> *It's more professional carer language, rather than patient language.*

### Recognizability

For both type 2 diabetes and dementia, recognition of the information on the website played an important role in the interview. When looking at the website, the participants assessed whether the information they found matched their experiences. This theme was most prominent among those looking at the pages about dementia. A number of participants said that, in their view, the information was too positive and did not give a realistic representation of the actual situation:

> *You realise that something is wrong because you have to urinate a lot. And very thirsty... that's all true. And other symptoms can arise. That's right enough.* [diabetes]

> *You keep talking about the GP here, but nowadays the diabetes nurse or the GP's medical assistant also do it.* [diabetes]

> *Lots of facts. But it also generates expectations that can't always be realised in practice. (dementia)*

> *You could also describe what the real situation is like. That you just say that this will sometimes not be possible.* [dementia]

### Care Route

The term *care route* meant little to most of the participants, especially to those who viewed the diabetes pages. It was more likely to be associated with different types of care providers than with different stages of the condition. However, some of the participants who viewed the pages about dementia did associate the term with the various stages of dementia:

> *I expected a route to the care institutions. Via the symptoms of dementia and how you get to the institutions.*

> *What I think of when I hear "care route", the name says it all...Distinguishing between early dementia and the middle and end phases.*

> *I know the route by bicycle. I think it's the same kind of route...but walking through the hospital for the care you're going to receive* [diabetes]

### Quality Information

Not many respondents spontaneously went to the quality information at the right-hand side of the diabetes page or within the *care route* of both conditions, and even fewer of them were really interested in it. When asked how they would choose a hospital, it turned out that most respondents made a choice based on distance, personal experience, and the GP's advice. Some participants did not realize that they could choose their own care provider, until we asked them to. They assumed the situation would dictate which care provider they would see:

> *Well, I think that I'd look at what's closest to me and ask the GP who they'd recommend if I don't know any of the doctors.*

> *You don't have a choice, you find yourself in a situation, depending on the treatment.*

> *And then I can see on the website what they can offer? Oh yes. I think that's pretty good.*

Most participants viewed the results of just 1 hospital, the closest hospital or the one they knew. Only a few participants saw that they could also select and compare multiple hospitals at once. When looking at the quality information, some participants did not realize that they could scroll down, as a result, failing to see information further down the page. Participants would have liked a link to the hospital's own website.

A number of participants had difficulty in understanding the quality information. This mainly concerned the numerical information (for instance, number of patients treated in the index year and percentage of patients who needed acute hospitalization), but some medical terms also needed more explaining. Some participants pointed out that it was important that the information was up to date. One participant also said that the *star rating* information (based on a client questionnaire score, the CQi) of the customer experience (for diabetes) was not very informative. They were more interested in reading the stories of previous patients in their own words:

> *It says here: The number of patients treated is 920, why is that important? Is the maximum 300 per physician or something like that? I don't know, no clue what this is supposed to mean.* [diabetes]

> *I'd be more interested in this...sure, what were the experiences of other people, but also: where was it measured, how many people…? What is Miletus?* [diabetes, referring to a quality programme run by Dutch insurers]

### Diabetes

A number of participants who viewed the Web page about diabetes felt that information was missing about the distance to the hospital and the waiting times. In addition, a variety of items were mentioned that people would like information about when choosing a hospital. For instance, about the number of examinations on a single day, whether it is possible to make an appointment with a diabetes nurse, the expertise of the specialists, and how much time a care provider has for a patient.

### Dementia

Informal carers of people with dementia felt that the most important aspects when choosing a hospital were personal attention, the content of the diagnosis consultation, the expertise of the medical specialists, and waiting times. However, rather than comparing hospitals, participants were more interested in comparing nursing homes, although that information was not available on these pages. For them, relevant aspects, besides information about the various types of living arrangements, were personal attention and quality of care in a nursing home. However, several of the informal carers said that information on the internet has virtually no influence on their choice of a nursing home. They would visit various homes and make a choice based on their own impressions and perceptions of the premises and its surroundings.

### *Usefulness*

The participants thought the information on KiesBeter was useful and helpful but not so much for them personally. They stated that the website is primarily suitable for informing and reassuring people who have just encountered the condition:

> *To let people know what type 2 diabetes is and how to deal with it.*

> *It's to give people reassurance. That there are enough options in the Netherlands for coping with it.*

> *I thought this information was very basic, for people who have just encountered the problem.*

The website made people think:

> *Sometimes losing weight is so effective that you don't need medicine anymore. That was news to me. I've never heard that.* [diabetes]

> *I thought it was useful to know all these numbers, stuff that you have to check daily.* [diabetes]

> *Interesting. It gets you to think and see whether we want a careplan, depending on how you define it* [dementia]

The participants thought the information was too general to apply to their own situation though. With regard to diabetes, it was the lifestyle information in particular that was too general, and for dementia, it was the information about how to relate to their relative:

> *Enough exercise, move about for half an hour. Well, "move about" is a little vague. What does that mean?* [diabetes]

> *Look here, it says you have to make sure your blood glucose doesn't get too high, but it doesn't say how.* [diabetes]

> *It doesn't say anywhere what to do if someone doesn't want to cooperate.* [dementia]

> *What it says is nice, but you've got to be able to do it yourself. It's more like suggestions, nothing very concrete.* [dementia]

## Discussion

### Overview

The aim of this study was to gain insight into how people process health and health care information on a website such as KiesBeter, especially the information concerning the *care route*, and to learn how a website like this could be improved. We had 38 participants think out loud when viewing the website, with verbal probing at specific moments. These participants were not *random* people, but most of them were familiar with the condition, being diabetes patient themselves or knowing someone with diabetes or being an informal carer for someone with dementia. We purposefully recruited participants familiar with these conditions and asked them before opening the Web page what they would want to know, so they would engage in a focused search. For this study, *care routes* were presented on test pages about diabetes and dementia but not available on the website. However, apart from the *care routes*, the pages were identical to the ones available on the website. Overall, 2 research questions were addressed in this study relating to information needs and information comprehension, including the *care route*. Cognitive interviews were held to better understand how people search for information and how they understand the information they find.

### Information Needs

Regarding health information, most participants were interested in the cause of the condition, the various forms, the way the condition progresses, and the consequences for daily life. However, the information on the website was too basic for many and was more appropriate for newly diagnosed patients. What the participants were looking for was more practical information that can be applied in day-to-day situations. Regarding health care information, no one expressed a need for quality, choice, or health care. The majority of respondents only looked at the available choice information after explicit prompting by the researcher. The finding that patients' own impression and perception often impact their choice of health care more than information on the Web is supported by Victoor et al [8].

### Information Comprehension

Before comprehending information, it has to be accessible. The first level of accessibility described by Eysenbach [3], physical accessibility, is not relevant here because we provided the participants with access to a computer and to the website. The other levels, findability, readability, and usability are included under this heading of information comprehension. *Recognizability* is a new theme that emerged from the analysis, which may be rooted in the fact that most participants were familiar with the condition they were reviewing. In our view, this reflects the reality that information about certain conditions will primarily be sought by patients with that condition. In that respect, *recognizability* is an aspect of usability, because for that group of consumers, the information is usable.

The *care route* is a newly developed presentation format showing the trajectory of a patient from diagnosis through various stages as the illness progresses, thereby, providing health information about the condition as well as health care information about different care providers. The added value of presenting information in the form of a *care route* is its scope: all available themes are visible at the same time as *stations* in the route, with the option to expand each station. For a focused search, it is immediately clear where to look for more detail, while it reminds users of other aspects of the condition they might otherwise overlook. That way, it helps users in finding information that they were not looking for (serendipity) and may trigger their curiosity, as is described by Pang et al [7].

The *care route* was received well as a format, although the name *care route* was found to be somewhat confusing. However, the information provided in the *care route* did not fit in very well with the actual situation and the information needs of the participants, partly because they were already familiar with the general information and felt that concrete and practical information, which they could apply to their personal situation, was missing. The participants liked that only a summary of the various stages was given first, so that they could then choose which stage they wanted to know more about. However, it was not clear to everyone that more information about the different stages as the illness progresses was available (that the bullets or *stations* in the *care route* could be expanded).

We found that a small majority of the participants (55%) were not familiar with the website KiesBeter and thought they would not be able to find it by themselves. Some also had problems finding specific items on the website because they were not where they expected them to be (findability/usability). The participants thought the information on the test pages was generally easy to understand and clearly presented. They were positive about the layout and design (usability). However, there were some concerns about the language and about the visibility of click-through options: the (medical) terms used could be explained better and the click-through options to additional information could be shown more clearly and consistently (readability). The website, with its short texts and click-through options for more information, was found to be suited for all kinds of visitors, those reading from top to bottom as well as those scanning the page for interesting items. However, for the participants, who have been coping with the condition for several years, most of the information was already known. Therefore, these pages are foremost seen as a good place to start for people who are only recently diagnosed with the condition (usability).

The quality information or choice information was not seen at all by some and not seen as relevant by most. Participants were not really interested in information about the quality of care providers, presented as a comparison between health care institutions. They made their choice based on distance, personal experience, and their GP's advice. This result corresponds with the findings by Victoor et al [8], who found that many patients make no active choice or choose a provider that is good enough based on only a few characteristics. In addition, we found that some of the information was poorly understood and the layout was not inviting enough for people to look at. The majority stated that they would sooner make a choice based on their own impressions than on information presented on the website.

### Limitations

The strength of our study lies in the richness of information we collected through the cognitive interviews. Most of our participants were able to engage in a focused search because they were already familiar with the conditions under study. A limitation was that the diversity within the group of participants was lower than expected. The people responding to our invitation turned out to be a selected group of, on average, highly educated, healthy, and confident internet users, as is often the case in-patient studies. Therefore, our results cannot be generalized to the general public. Future work should include a wider diversity of participants. Another limitation of our study, like many other studies on choice, for instance, the Zwijnenberg study [22], is that participants did not need to make a choice for a health care provider. The choice was hypothetical and therefore less relevant. It is unclear how people would choose in real situations. Nevertheless, we believe our results contribute to the knowledge on the design and the user experience of consumer health websites.

### Comparison With Other Studies

The themes resulting from the analysis closely resembled the levels described by Eysenbach [4]. This study confirms that information on a website should be kept as simple as possible [12,13,16]. The results about choice information are in line with the results Hibbart and Gutacker and Damman et al presented [11,12]. Zwijnenberg et al [22] stress the need for flexible, user-friendly websites or *information on demand*. With the presentation of the *care route* with all stages of the condition available in 1 overview and the option to expand each station for more information, the developers have aimed at just that. Furthermore, people told us they chose a hospital based on distance, personal experience, and their GP's advice. This is also in line with the results by Victoor et al [8]. Zwijnenberg et al [22] explored patients' preferences regarding the way comparative information is presented and the value of tailoring information to specific groups, but their group of participants, like ours, was not diverse enough to lead to a conclusion.

### Implications for Practice

The health information on the reviewed Web pages was found to be generally easy to understand and clearly presented, but the layout of the pages could be improved. For visitors familiar with the condition on the Web page, the information was too general. They would prefer to find more specific information that is applicable to their daily life.

One of the aims of presenting quality information is to encourage people to choose a care provider by comparing health care institutions, but we found that most people were not inclined to do so but rather would make a choice based on distance, personal experience, and the GP's advice. Maybe focusing on choice is not the best possible approach when the ultimate goal is accessible and affordable health care. Maybe providing relevant and usable health information in itself will guide consumers' choices enough to move toward that goal.

Our findings apply to the KiesBeter website but are relevant for other consumer websites as well. We conclude that it is important to involve future users in designing a health care information website, for instance, by conducting cognitive interviews as we did. This may help improve the quality and the usability of the website, preferably before it is implemented.

### Conclusions

The cognitive interviews gave numerous insights into how information on a website such as KiesBeter is processed and understood. The study indicates that for these respondents, who have been coping with the condition for several years, the added value of KiesBeter is small. Their impression is that it is more appropriate for people who have only recently been diagnosed with a condition and are looking for basic information about that condition. Providing choice information on a website does not seem to influence the way people make choices concerning a hospital. In addition, it seems not to encourage people to make an informed choice for a care provider. Finally, we observed that the name *care route* is not clear to everyone. On the other hand, the *care route* offers a clear overview of the various stages as the condition progresses.


## Conflicts of Interest

None declared.


## Multimedia Appendix 1

Screenshot of KiesBeter, January 2014, Diabetes.

[PDF File (Adobe PDF File), 133KB - jmir_v21i7e11340_app1.pdf ]

## References


1. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality-of-care information? A systematic review. Med Care 2009 Jan;47(1):1-8. [doi: 10.1097/MLR.0b013e3181808bb5] [Medline: 19106724]
2. National Health Service. URL: https://www.nhs.uk/ [accessed 2019-05-21] [WebCite Cache ID 78XFCxxZX]
3. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google is here to stay but health care professionals are still valued: an analysis of health care consumers' internet navigation support preferences. J Med Internet Res 2017 Dec 14;19(6):e210 [FREE Full text] [doi: 10.2196/jmir.7489] [Medline: 28615156]
4. Eysenbach G. Designevaluation of consumer health information on web sites. In: Consumer Health Informatics: Informing ConsumersImproving Health Care. New York: Springer-Verlag; 2005:34-60.
5. Lambert SD, Loiselle CG, Macdonald ME. An in-depth exploration of information-seeking behavior among individuals with cancer: part 1: understanding differential patterns of active information seeking. Cancer Nurs 2009;32(1):11-23. [doi: 10.1097/01.NCC.0000343372.24517.bd] [Medline: 19104197]
6. Higgins O, Sixsmith J, Barry MM, Domegan C. European Centre for Disease Prevention and Contro. 2011. A literature review on health information-seeking behaviour on the web: a health consumer and health professional perspective URL: https://tinyurl.com/y2fxaawk [accessed 2019-06-04] [WebCite Cache ID 78s99DiL4]
7. Pang PC, Chang S, Verspoor K, Pearce J. Designing Health Websites Based on Users' Web-Based Information-Seeking Behaviors: A Mixed-Method Observational Study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
8. Victoor A, Delnoij DM, Friele RD, Rademakers JJ. Determinants of patient choice of healthcare providers: a scoping review. BMC Health Serv Res 2012 Aug 22;12:272 [FREE Full text] [doi: 10.1186/1472-6963-12-272] [Medline: 22913549]
9. Edgman-Levitan S, Cleary PD. What information do consumers want and need? Health Aff (Millwood) 1996;15(4):42-56. [doi: 10.1377/hlthaff.15.4.42] [Medline: 8991253]
10. Hibbard JH, Jewett JJ. Will quality report cards help consumers? Health Aff (Millwood) 1997;16(3):218-228. [doi: 10.1377/hlthaff.16.3.218] [Medline: 9141339]

11. Hibbard JH, Slovic P, Jewett J. Informing consumer decisions in health care: implications from decision-making research. Milbank Q 1997;75(3):395-414 [FREE Full text] [Medline: 9290635]
12. Damman OC, Hendriks M, Rademakers J, Delnoij D, Groenewegen P. How do healthcare consumers process and evaluate comprative healthcare information? A qualitative study using cognitive interviews. BMC Public Health 2009 Nov 20;9:423 [FREE Full text] [doi: 10.1186/1471-2458-9-423] [Medline: 19930564]
13. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003 Jan;24:413-433. [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]
14. Bettman JR, Kakkar P. Effects of information presentation format on consumer information acquisition strategies. J Consum Res 1977;3(4):233-240.
15. Slovic P. Towards understanding and improving decisions. In: Human PerformanceProductivity. Information ProcessingDecision Making. United Kingdom: Psychology Press; 1982:2.
16. Ruijter L, Hendriks M, Malanda U, Boer de D. Visual care route as a way to present choice information about care: report of literature review and interviews with patient organizations. Utrecht: Nivel; 2017.
17. Ossebaard HC, Seydel ER, van Gemert-Pijnen L. Online usability and patients with long-term conditions: a mixed-methods approach. Int J Med Inform 2012 Jun;81(6):374-387. [doi: 10.1016/j.ijmedinf.2011.12.010] [Medline: 22261086]
18. [Promote quality and clarity of care]. Zorginstituut Nederland URL: https://www.zorginstituutnederland.nl/over-ons/taken/bevorderen-kwaliteit-en-inzichtelijkheid-van-de-zorg [accessed 2019-05-21] [WebCite Cache ID 78XGnyvLB]
19. KiesBeter. URL: https://www.kiesbeter.nl/ [accessed 2019-06-04] [WebCite Cache ID 78tVnDMD6]
20. European Pathway Association. CARE PATHWAYS URL: http://e-p-a.org/care-pathways/ [accessed 2019-05-21] [WebCite Cache ID 78XGsGQNA]
21. Boeije HR. [Analyzing in qualitative research: thinking and doing]. Amsterdam: Boom; 2005.
22. Zwijnenberg NC, Hendriks M, Bloemendal E, Damman OC, de Jong JD, Delnoij DM, et al. Patients' need for tailored comparative health care information: a qualitative study on choosing a hospital. J Med Internet Res 2016 Nov 28;18(11):e297 [FREE Full text] [doi: 10.2196/jmir.4436] [Medline: 27895006]


## Abbreviations

**DVN:** Dutch Diabetes Association
**NIVEL:** Netherlands Institute for Health Services Research


*Edited by G Eysenbach; submitted 20.07.18; peer-reviewed by A Marcu, PCI Pang, E Ray Chaudhuri; comments to author 06.10.18; revised version received 29.01.19; accepted 07.04.19; published 08.07.19.*

*Please cite as:*
*Wiegers TA, Hendriks M, Malanda U, de Boer D*
*Users' Experiences With Web-Based Health Care Information: Qualitative Study About Diabetes and Dementia Information Presented on a Governmental Website*
*J Med Internet Res 2019;21(7):e11340*
*URL: https://www.jmir.org/2019/7/e11340/*
*doi:10.2196/11340*
*PMID:31287066*

Viewpoint

# Hacking 9-1-1: Infrastructure Vulnerabilities and Attack Vectors

Mat Goebel[1], MD, MAS, NR-P; Christian Dameff[2], MD; Jeffrey Tully[3], MD

[1]Department of Emergency Medicine, University of Massachusetts Medical School–Baystate, Springfield, MA, United States

[2]Department of Emergency Medicine, University of California San Diego, San Diego, CA, United States

[3]Department of Anesthesiology and Pain Medicine, UC Davis Medical Center, Sacramento, CA, United States

**Corresponding Author:**
Mat Goebel, MD, MAS, NR-P
Department of Emergency Medicine
University of Massachusetts Medical School–Baystate
759 Chestnut St
Springfield, MA,
United States
Phone: 1 413 794 0000
Email: mgoebel@gmail.com

## Abstract

9-1-1 call centers are a critical component of prehospital care: they accept emergency calls, dispatch field responders such as emergency medical services, and provide callers with emergency medical instructions before their arrival. The aim of this study was to describe the technical structure of the 9-1-1 call-taking system and to describe its vulnerabilities that could lead to compromised patient care. 9-1-1 calls answered from mobile phones and landlines use a variety of technologies to provide information about caller location and other information. These interconnected technologies create potential cyber vulnerabilities. A variety of attacks could be carried out on 9-1-1 infrastructure to various ends. Attackers could target individuals, groups, or entire municipalities. These attacks could result in anything from a nuisance to increased loss of life in a physical attack to worse overall outcomes owing to delays in care for time-sensitive conditions. Evolving 9-1-1 systems are increasingly connected and dependent on network technology. As implications of cybersecurity vulnerabilities loom large, future research should examine methods of hardening the 9-1-1 system against attack.

***(J Med Internet Res 2019;21(7):e14383)*** doi:10.2196/14383

**KEYWORDS**

cybersecurity; emergency medical services; emergency medical dispatch; emergency medical service communication systems

## Introduction

9-1-1 call centers are a critical component of prehospital care: they accept emergency calls, dispatch field responders such as emergency medical services (EMS) units, and provide callers with emergency medical instructions before their arrival. The catalyst for creating the 9-1-1 system in the United States came in 1957, when the National Association of Fire Chiefs recommended using a single number to report fires. In 1967, the Federal Communications Commission worked with the American Telephone and Telegraph Company to establish a single emergency number [1]. 9-1-1 was selected because it had never been authorized as an office code, area code, or service code. The first 9-1-1 call was made on February 16, 1968. Later, Congress passed a legislation making 9-1-1 the official emergency number throughout the United States.

Today, emergencies can be reported through 9-1-1 from 96% of the geography of the United States, placing an estimated 240 million calls annually. There are 5783 public safety answering points (PSAPs, colloquially known as 9-1-1 call centers) in 3135 counties throughout the United States [2]. The public relies on the 9-1-1 service to get assistance during their time of need. Similarly, municipalities rely on PSAPs to know what manner of emergencies are occurring where and when, so that appropriate resources can be dispatched—fire services, emergency medical services, and police services. Medical emergencies comprise 64% of emergency calls [3].

Although the original implementation of basic 9-1-1 services relied mostly on telephone technology, enhanced 9-1-1 (E9-1-1), deployed in the 1980s, introduced significant new elements aimed at identifying the caller's telephone number and location. Around the same time, new tools were developed to assist dispatchers in handling emergency calls: advanced hardware and software to receive and display the caller's number and location data, computer-aided dispatch (CAD) systems to assign and track field responders, and mapping software to display both caller and responder locations. These tools greatly enhanced emergency response capabilities but required additional computer, network, and server infrastructure beyond the rudimentary elements of the telephone network. As capabilities further expanded, systems that were once safely within the walled garden of the telephone company were increasingly connected to less well-regulated networks such as the public internet. Though undoubtedly valuable from a patient outcome perspective, this also introduced new risks.

According to the Department of Homeland Security, the emergency services sector will likely become more of a target for cyberattacks as systems become more interconnected and dependent on information technology for daily operations [4]. Disruption or unavailability of the 9-1-1 service owing to cyberattacks targeting either data or physical infrastructure has occurred and could delay response to emergencies, potentially threatening life, limb, and property [5]. In this study, we aimed to describe the technical structure of the 9-1-1 call-taking system and describe its vulnerabilities that could lead to compromised patient care.

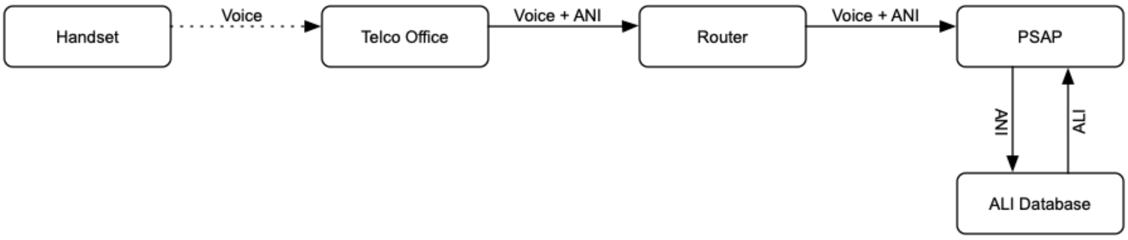
## Technical Infrastructure

### Landline Emergency Calls

An emergency call placed from a landline is the simplest scenario for emergency call taking (see Figure 1). When the call is placed, the telephone company office (telco) attaches the telephone number to the call. This is called Automatic Number Identification (ANI). The telco sends the call to a router, which forwards the call to the appropriate PSAP (also known as a 9-1-1 call center) based on the ANI. If for some reason the data containing the ANI are corrupted or the call arrives without an ANI, the call is routed to a default PSAP.

The call is answered at the PSAP by a dispatcher, often with the help of a CAD system. One of the dispatcher’s first priorities is to determine the location of the call. The CAD system uses the ANI to search a database, which returns Automatic Location Information (ALI). The ALI record contains the address, phone number, and which services are assigned to the caller’s position.

**Figure 1.** Schema for wired 9-1-1 phone call. ANI: automatic number identification; ALI: automatic location information; PSAP: public safety answering point.




### Wireless Phase 1 Emergency Calls

A wireless emergency call’s routing varies depending on the E9-1-1 implementation. In a phase 1 implementation, the call is sent from a cell tower to a mobile switching center (see Figure 2). The mobile switching center sends the call and cell tower information to a mobile positioning center. The positioning center returns a pseudo-ANI (pANI), which corresponds to the telephone number of the cell site. The positioning center also creates temporary location information (using the cell tower location) that can be looked up by the PSAP. The switching center attaches the pANI to the call and forwards the call to a router, which selects the appropriate PSAP. The PSAP then looks up the location information, which returns the position approximated by the location of the cell tower.

### Wireless Phase 2 Emergency Calls

In a phase 2 implementation, additional equipment determines the caller’s location more precisely (see Figure 3). The exact equipment and methods vary between cellular providers, but the end relay of position information is the same. When a call is placed, the same chain of events occurs as in phase 1. However, when the mobile positioning center receives the cell tower information, it requests the caller’s location from position determination equipment, which returns latitude and longitude reported by the mobile device. The positioning center returns a pANI as before but enters the actual latitude and longitude of the call in the location database rather than an estimated position based on cell tower information. The switching center sends the call to a router, which selects the appropriate PSAP. When the PSAP queries the location database, the ALI includes the latitude and longitude of the call. In this implementation, a mapping or other geographic information system is commonly integrated with the CAD to automatically plot the call’s location.

**Figure 2.** Schema for a phase 1 wireless 9-1-1 call. pANI: pseudo automatic number identification; ALI: automatic location information; PSAP: public safety answering point; CBN: call back number.

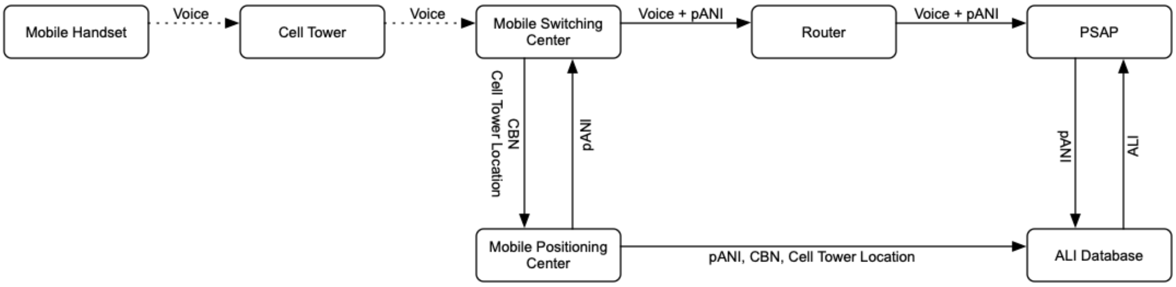
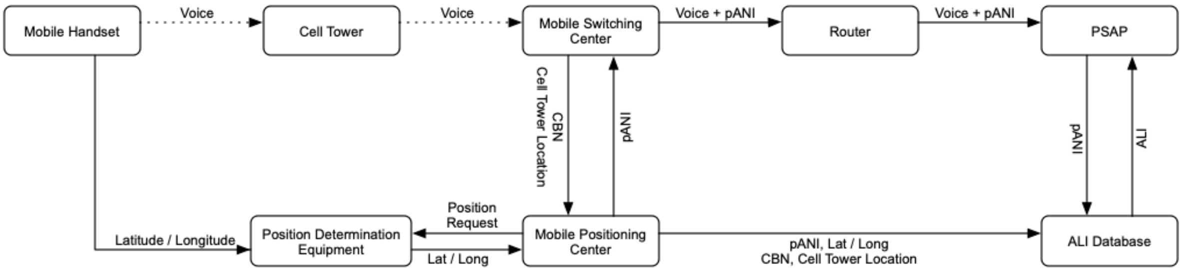


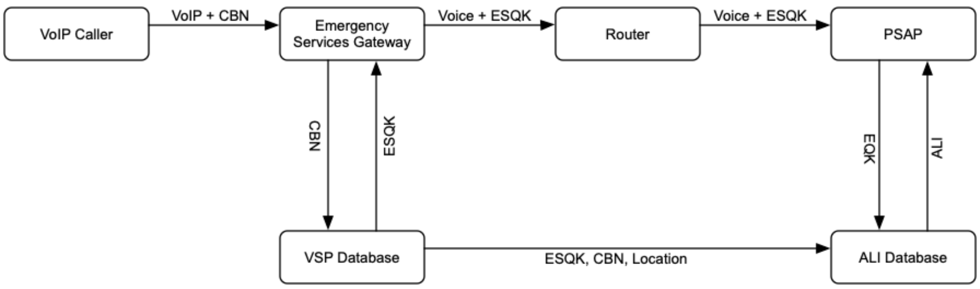

**Figure 3.** Schema for a phase 2 wireless 9-1-1 call. pANI: pseudo automatic number identification; ALI: automatic location information; PSAP: public safety answering point; Lat: latitude; Long: longitude; CBN: call back number.




## Voice-Over-IP Emergency Calls

Voice-over-IP (VoIP), or voice communications delivered by internet-connected networks from providers such as Skype (Skype Technologies, Palo Alto, CA) and Vonage (Vonage Holdings Corp, Holmdel, New Jersey), presents a special problem for 9-1-1 location services (see Figure 4). In this scenario, the VoIP service provider (VSP) maintains a database that associates their subscribers with the appropriate PSAP and location information. These data are usually based on billing information provided by the user. When an emergency call is placed, the call is sent to an Emergency Services Gateway. The gateway searches the VSP database, returning the correct PSAP. The VSP database uses the subscriber's information to create a temporary location database entry. The call is sent from the gateway to a router, which directs the call to the appropriate PSAP. When the PSAP retrieves location information from the database, it receives the location information provided by the VSP database.

**Figure 4.** Schema for a VoIP call. VoIP: voice-over-internet-protocol; ESQK: emergency services query key; CBN: call back number; ALI: automated location information; PSAP: public safety answering point; VSP: voice-over-IP service provider.



## Vulnerabilities

Information security vulnerabilities fall under 3 main categories: confidentiality, integrity, and availability—often referred to as the CIA triad.

### Breaches of Confidentiality

An attack on confidentiality results in the disclosure of previously private information or communications to one or more parties who are not authorized to receive or view it. Surveillance of the 9-1-1 system could provide attackers with information valuable to planning an incident. Metrics such as call volumes and response times could be used to maximize damage during a physical attack or increase the likelihood of success of criminal acts.

### Breaches of Integrity

An attack breaching integrity violates the correct functioning of a system, producing aberrant results desired by the attacker. An example of this is an attack called swatting, where an individual misdirects law enforcement resources to target another individual, often for personal reasons such as revenge. A December 2017 swatting incident in Wichita, Kansas, led to the shooting of an unarmed man, resulting in his death [6]. Swatting can occur via several methods.

Teletypewriter (TTY) services are intended as communication relays between deaf and hearing persons. A deaf person connects to the relay service using a special terminal with a keyboard or an online service, then requests to an operator who they would like to speak with. The operator connects the call, then verbatim reads messages from the terminal to the caller, and types messages from the caller to the terminal. If an attacker connects to a TTY service and describes an emergency to the operator, the operator will call 9-1-1 and report exactly what the TTY user is typing. Using a TTY service removes some of the attacker's potentially identifying information (such as their voice). In addition, TTY relay services are required to ensure user confidentiality and are prohibited from keeping records of the contents of any conversation [7]. This method was used to swat Ashton Kutcher and Justin Bieber in 2013 [8].

Other integrity attacks include spoofing apps or services that generate false location or caller identification (ID) information by modifying the ANI or ALI record. This method will not work with a landline because the ANI/ALI is not dynamic. Many spoofing services treat emergency calls differently and will replace the correct ANI data before connecting the call. In addition, when an emergency call is placed from a spoofing service, the service provider ignores any ANI information received from the caller and inserts the correct data. In some rare circumstances, VoIP providers have been found to forward user-provided ANI data without verification or alteration, which could allow spoofing. In addition, this information can seemingly be updated without verification.

If an attacker successfully spoofed ANI information, they could circumvent automatic PSAP routing to target a specific PSAP. The routers that forward emergency calls have a direct dial phone number that is used to transfer calls between PSAPs when a caller is incorrectly routed or if fire and police services are dispatched separately. Although the direct dial numbers for PSAPs' 9-1-1 network connections and 10-digit emergency lines are kept secret (and are supposed to be nondialable), they can sometimes be discovered during a 9-1-1 call that involves a transfer if the PSAP equipment is not configured to mute the dialing tones used to connect the call. If an attacker utilized a spoofed ANI to call the router at the newly discovered direct dial number, they could place a seemingly authentic call to a specific PSAP. Targeting a specific PSAP could improve an attacker's chances of getting the desired response to a call, such as a swatting attack.

All location determination mechanisms rely on the PSAP querying the ALI database. If an attacker altered this database, they could change the ALI information for any phone number to their target's address or change the address associated with their target's phone number. If an attacker denied access to the ALI database, dispatchers would rely entirely on the information provided by the caller. It is notable that the ALI database represents a single point of failure in every location-determining scheme.

The VSP database that maintains the mapping between the subscriber and location could also be targeted. Modification of the database or a denial of service attack would have the same effect as an attack on the ALI database, but may be secured differently.

Mobile handsets without an active service plan can still be used to place 9-1-1 calls. These nonservice initiated (NSI) devices do not provide a call back number because they do not have a service plan. Instead, they provide a phone number with area code 9-1-1 with the last 7 digits of the device's electronic serial number or international mobile equipment identity number in place of a phone number. Such a device could be purchased with cash and used to place a call, removing the standard information links between the 9-1-1 call and caller.

Wireless phase 2 calls utilize the location data from the mobile handset itself. It is possible to inject arbitrary coordinates into the handset, either by global positioning system spoofing or modifying the device firmware such that it reports prespecified coordinates. The call will be routed to the appropriate PSAP based on the cell tower location but the ALI record will contain the arbitrary location which is then displayed at the PSAP. This is not foolproof, however, as most cellular providers also use cell tower triangulation as part of the location-determination process.

Attacking the location-determination systems (as described above) or denying service to the entire 9-1-1 infrastructure could delay response to an incident to increase collateral damage. Modifying the VSP or ALI records for an individual could lead to emergency response being directed to the wrong location, delaying emergency services to a targeted individual.

Another integrity attack could target first responders or enhance damage from a physical attack. By misdirecting resources throughout a municipality, an attacker could delay emergency response to a planned physical attack, increasing collateral damage. In addition, an attacker could create a major incident that concentrates emergency responders in a specific location for the purpose of directly attacking responders. These 2 types of attack could also be used in combination. Altering or denying service to the VSP or ALI database could increase the credibility of calls used to initiate such incidents.

### Breaches of Availability

An attack on availability would make a PSAP unavailable to the public. This could be accomplished by a telephone denial of service attack, as occurred during March 2016 in Phoenix. An attacker released a malicious script via a link on Twitter that caused devices to repeatedly call the PSAP without allowing the user to hang up. This produced thousands of false 9-1-1 calls, filling up the phone lines and preventing the public from reporting emergencies [6]. The Federal Bureau of Investigation and Department of Homeland Security have noted an increase in denial of service attacks accomplished by inundating a PSAP's phone lines with robot dialing [4]. In a December 2018 incident, though not a cyberattack, a failure in the network of internet service provider, CenturyLink, made various PSAPs unreachable in Washington state, Oregon, Minnesota, Massachusetts, Idaho, New Mexico, Missouri, Arizona, and Colorado [9].

Inundating a PSAP with phone calls can make the phone system unavailable. The same attack can be applied to computer systems with the internet or network traffic in a distributed denial of service (DDos) attack. A March 2015 DDos attack on the City of Madison, Wisconsin, targeted municipal websites with the unintended consequence of disrupting 9-1-1 infrastructure. In March 2018, a ransomware attack (malicious software that renders a computer unusable unless an attacker is paid to unlock the system) made the City of Baltimore's CAD system unavailable to dispatchers for almost 24 hours while systems were restored. Although the PSAP was still reachable, call takers had to gather information by hand that would have normally been displayed automatically, slowing the dispatching process [9].

## Discussion

### Effects on Patient Care

All 9-1-1–focused attacks to date have been linked to criminal hackers rather than state-sponsored actors [4]. However, as a form of terrorism or even warfare, attacks on the 9-1-1 system could incite panic and loss of life to benefit any enemy of the state. Cyber salvos against the Ukrainian power grid in December 2015, by actors believed to be affiliated with the Russian government, serve as an example of an operation in which an emergency response system may be compromised with failure of underlying infrastructure [10]. Although the direct effects of 9-1-1 attacks on patient care have not been studied, we can infer from other data that delays in emergency care worsens outcomes [11]. Any number of time-sensitive conditions (heart attack, stroke, sepsis, and trauma) could have delayed response times during a 9-1-1 attack, leading to worsened morbidity and mortality.

### Next Generation Attacks

Currently, there is a national effort to upgrade traditional telephone-based 9-1-1 architecture to Next Generation 9-1-1 (NG 9-1-1). This is a collection of technologies that allow PSAPs to receive short message service text messages, images, and even live video streams in addition to existing landline, mobile, and VoIP calls. NG 9-1-1 promises increased capabilities for precisely locating callers and responding to mass casualty incidents, infrastructure disruptions, and natural disasters. However, converting to a system dependent on internet-connected networks raises the possibility of inheriting the same cybersecurity vulnerabilities that plague existing connected infrastructure while continuing to be susceptible to the threat models affecting telephone-based systems that we have already discussed.

The National Public Safety Telecommunications Council has outlined the future of EMS communications [5]. Software applications that allow rapid sharing of data between the field and the hospital reduce time to treatment in heart attack and stroke. Sensor networks in vehicles will detect collisions and wearable health devices will detect cardiac arrest to activate 9-1-1 automatically. Prehospital agencies will share telemetry and live video, allowing for physician consultation in the field. Hospitals will have virtual dashboards that give them the status of arriving patients. Although these technologies will be a boon for patient care, adding cyber infrastructure and network interconnection may introduce a variety of vulnerabilities.

Fortunately, centers upgrading to NG 9-1-1 have the potential to mitigate cybersecurity vulnerabilities and related disruptions. Important best practices include frequently patching vulnerable software applications, segmenting networks to protect infrastructure from a direct connection to the internet, and developing incident response plans and teams to react to cybersecurity events that occur [12].

### Future Directions for Readiness

A variety of attacks could be carried out on the 9-1-1 infrastructure to various ends. Attackers could target individuals, groups, or entire municipalities. These attacks could result in anything from a nuisance to increased loss of life in a physical attack to worse overall outcomes owing to delays in care for time-sensitive conditions. Evolving 9-1-1 systems are increasingly connected and dependent on network technology. As implications of cybersecurity vulnerabilities loom large, future research should examine methods of hardening the 9-1-1 system against attack—whether through the technical work of designing and developing increasingly secure software platforms utilizing next generation encryption and authentication methodology to further increase trust of network communications, through implementing regulatory frameworks with incentives for procurement and operational guidance, or through, perhaps most feasible in the short term, raising awareness among individual and regional centers of existing security challenges and corresponding cybersecurity best practices.

## Conflicts of Interest

None declared.

## References

1. Dayharsh TI, Yung TJ, Hunter DK, Ivy SC. Update on the national emergency number 911. Ieee T Veh Technol 1979 Nov;28(4):292-297. [doi: 10.1109/T-VT.1979.23804]
2. National Emergency Number Association. 2018. 9-1-1 Statistics URL: https://www.nena.org/page/911Statistics [accessed 2018-12-29] [WebCite Cache ID 77TIbS6Sq]
3. Evarts B. National Fire Protection Association. 2018 Oct. Fire Loss in the United States During 2017 URL: https://www.nfpa.org/~/media/FD0144A044C84FC5BAF90C05C04890B7.ashx
4. US Department of Homeland Security. Public Intelligence. 2016. DHS Assessment: Cyber Targeting of the US Emergency Services Sector Limited, But Persistent URL: https://publicintelligence.net/dhs-cyber-targeting-ess/
5. National Association of State Chief Information Officers. National Association of State Chief Information Officers. 2016 Apr. Cyber Disruption Response Planning Guide URL: https://www.nascio.org/Publications/ArtMID/485/ArticleID/358/Cyber-Disruption-Response-Planning-Guide
6. Lee D. BBC World News. 2016 Oct 31. Teen Hacker 'Accidentally' Causes 911 Attack URL: https://www.bbc.com/news/technology-37821862 [accessed 2019-01-15] [WebCite Cache ID 77TIbEHIG]
7. Federal Communications Commission. 2011. Telecommunications Relay Service - TRS URL: https://www.fcc.gov/consumers/guides/telecommunications-relay-service-trs [accessed 2018-11-19]
8. Pepitone J. CNN Money. 2013 Apr 14. 'Swatting' Celebrities is Far Too Simple URL: https://money.cnn.com/2013/04/14/technology/security/swatting-caller-id/index.html [accessed 2018-11-19] [WebCite Cache ID 77cDF07Tk]
9. Cimpanu C. ZDNet. 2018 Dec 28. CenturyLink Outage Takes Down Several 911 Emergency Services Across the US Internet URL: https://www.zdnet.com/article/centurylink-outage-takes-down-several-911-emergency-services-across-the-us/ [accessed 2019-01-15] [WebCite Cache ID 77TIaeyiv]
10. Finkle J. Reuters. 2016 Jan 8. US Firm Blames Russian 'Sandworm' Hackers for Ukraine Outage URL: https://www.reuters.com/article/us-ukraine-cybersecurity-sandworm-idUSKBN0UM00N20160108 [accessed 2019-05-11] [WebCite Cache ID 78J09Ztf0]
11. Jena AB, Mann NC, Wedlund LN, Olenski A. Delays in emergency care and mortality during major US marathons. N Engl J Med 2017;376(15):1441-1450 [FREE Full text] [doi: 10.1056/NEJMsa1614073] [Medline: 28402772]
12. The Department of Homeland Security. 2018 Nov. Office of Emergency Communications: Cyber Risks to Next Generation 9-1-1 URL: https://www.dhs.gov/sites/default/files/publications/Cybersecurity%20Risks%20for%20NG9-1-1%20%28100418%29_508C_FINAL.pdf

## Abbreviations

**ALI:** Automatic Location Information
**ANI:** Automatic Number Identification
**CAD:** computer-aided dispatch
**DDos:** distributed denial of service
**E9-1-1:** enhanced 9-1-1
**NG 9-1-1:** Next Generation 9-1-1
**pANI:** pseudo-ANI
**PSAP:** public safety answering point
**telco:** telephone company office
**TTY:** teletypewriter
**VoIP:** Voice-over-IP
**VSP:** VoIP service provider

*Edited by E Perakslis; submitted 14.04.19; peer-reviewed by B Woods, M Stanley; comments to author 08.05.19; revised version received 14.05.19; accepted 21.05.19; published 09.07.19.*

*Please cite as:*
*Goebel M, Dameff C, Tully J*
*Hacking 9-1-1: Infrastructure Vulnerabilities and Attack Vectors*
*J Med Internet Res 2019;21(7):e14383*
*URL: https://www.jmir.org/2019/7/e14383/*
*doi:10.2196/14383*
*PMID:31290401*

Review

# Impact of Clinicians' Use of Electronic Knowledge Resources on Clinical and Learning Outcomes: Systematic Review and Meta-Analysis

Lauren A Maggio[1], PhD; Christopher A Aakre[2], MD; Guilherme Del Fiol[3], MD, PhD; Jane Shellum[4], MHA; David A Cook[2], MD

[1]Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, United States

[2]Division of General Internal Medicine, Mayo Clinic College of Medicine and Science, Rochester, MN, United States

[3]Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, United States

[4]Center for Translational Informatics and Knowledge Management, Mayo Clinic, Rochester, MN, United States

**Corresponding Author:**
Lauren A Maggio, PhD
Department of Medicine
Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD, 20814-4799
United States
Phone: 1 301 295 4371
Email: lauren.maggio@usuhs.edu

## Abstract

**Background:** Clinicians use electronic knowledge resources, such as Micromedex, UpToDate, and Wikipedia, to deliver evidence-based care and engage in point-of-care learning. Despite this use in clinical practice, their impact on patient care and learning outcomes is incompletely understood. A comprehensive synthesis of available evidence regarding the effectiveness of electronic knowledge resources would guide clinicians, health care system administrators, medical educators, and informaticians in making evidence-based decisions about their purchase, implementation, and use.

**Objective:** The aim of this review is to quantify the impact of electronic knowledge resources on clinical and learning outcomes.

**Methods:** We searched MEDLINE, Embase, PsycINFO, and the Cochrane Library for articles published from 1991 to 2017. Two authors independently screened studies for inclusion and extracted outcomes related to knowledge, skills, attitudes, behaviors, patient effects, and cost. We used random-effects meta-analysis to pool standardized mean differences (SMDs) across studies.

**Results:** Of 10,811 studies screened, we identified 25 eligible studies published between 2003 and 2016. A total of 5 studies were randomized trials, 22 involved physicians in practice or training, and 10 reported potential conflicts of interest. A total of 15 studies compared electronic knowledge resources with no intervention. Of these, 7 reported clinician behaviors, with a pooled SMD of 0.47 (95% CI 0.27 to 0.67; $P$<.001), and 8 reported objective patient effects with a pooled SMD of 0.19 (95% CI 0.07 to 0.32; $P$=.003). Heterogeneity was large ($I^2$>50%) across studies. When compared with other resources—7 studies, not amenable to meta-analytic pooling—the use of electronic knowledge resources was associated with increased frequency of answering questions and perceived benefits on patient care, with variable impact on time to find an answer. A total of 2 studies compared different implementations of the same electronic knowledge resource.

**Conclusions:** Use of electronic knowledge resources is associated with a positive impact on clinician behaviors and patient effects. We found statistically significant associations between the use of electronic knowledge resources and improved clinician behaviors and patient effects. When compared with other resources, the use of electronic knowledge resources was associated with increased success in answering clinical questions, with variable impact on speed. Comparisons of different implementation strategies of the same electronic knowledge resource suggest that there are benefits from allowing clinicians to choose to access the resource, versus automated display of resource information, and from integrating patient-specific information. A total of 4 studies compared different commercial electronic knowledge resources, with variable results. Resource implementation strategies can significantly influence outcomes but few studies have examined such factors.

*(J Med Internet Res 2019;21(7):e13315)* doi:10.2196/13315

**KEYWORDS**

medical education; information systems; educational technology; clinical decision support; health information technology

## Introduction

Clinicians and trainees frequently identify clinical questions while caring for patients [1]. They have been trained, and often attempt, to answer these questions using a variety of resources, including increasing use of electronic resources [2-4]. Electronic knowledge resources have been defined as “electronic (computer-based) resources comprising distilled (synthesized) or curated information that allows clinicians to select content germane to a specific patient to facilitate medical decision making” [5]. Commonly used electronic knowledge resources include commercial products, such as UpToDate, Micromedex, and Epocrates [6,7]; locally developed products, such as McMaster Premium LiteratUre Service (PLUS) [8]; and crowdsourced resources, such as Wikipedia [9]. Electronic knowledge resources are related to, but distinct from, decision-support tools that provide pop-up alerts, reminders, and other push notifications or databases of unsynthesized information, such as MEDLINE.

Electronic knowledge resources are commonly used in clinical practice and typically require significant resources, including the financial investment in procuring access and clinicians' investment of time in learning to use them [10]. However, their impact on patient care and learning outcomes is incompletely understood [4,11]. Previous reviews of health information resources have, in general, broadly focused on clinical decision-support tools [12,13]. One review characterized features of clinical information retrieval technology that promote its use [14] but did not examine the specific knowledge resources themselves. Another review of clinicians' information-seeking behaviors identified textbooks, colleagues, journal articles, professional websites, and medical libraries as information sources but did not report the outcomes associated with using these sources [15]. A review of clinical questions noted the use of knowledge resources to answer such questions but did not directly address knowledge resources [1]. Moreover, the age of these reviews (ie, the most recent having been published in 2014) limits their application to current practice. An up-to-date, comprehensive synthesis of evidence regarding the effectiveness of electronic knowledge resources could guide clinicians, health care system administrators, medical educators, and informaticians in making evidence-based decisions about their purchase, implementation, and use. Thus, we conducted a systematic review to answer the following question: What is the impact of electronic knowledge resources for clinicians on clinical and learning outcomes?

## Methods

This study is part of a large systematic review of knowledge resources and point-of-care learning that was planned, conducted, and reported in adherence to standards of quality for reporting meta-analyses [16].

### Search Strategy and Study Selection

With support from an experienced reference librarian, on February 14, 2017, we simultaneously searched MEDLINE, Embase, PsycINFO, and the Cochrane Library Database using Ovid’s integrated search interface for comparative studies of electronic knowledge resources. We used the databases’ controlled vocabulary thesauri, Web searches, the research teams' files, and previous reviews [1,6,13,14,17] to create and refine the search strategy and supplemented the database search by examining the full bibliography of these reviews. Search terms included a combination of keywords and controlled vocabulary terms (eg, *information-seeking behavior*, *point-of-care systems*, *drug information services*, *UpToDate*, and *Micromedex*). Multimedia Appendix 1 describes the complete search strategy. We limited our search to studies published after January 1, 1991, the year in which the World Wide Web was first described. We made no exclusions based on language.

### Article Selection

We included all original, comparative studies that evaluated clinicians' use of an electronic knowledge resource, using quantitative outcomes of knowledge, skills in a test setting, attitudes, behaviors with real patients, patient effects, and costs. We required that outcome measures relate to a clinical decision for a specific patient or clinical vignette; we excluded studies measuring only general experiences or overall perceived impact. Measurements in a test setting had to be objectively assessed, as opposed to clinician-reported, and performed without immediate support from the knowledge resource (ie, evaluating sustained impact on knowledge after a period of access, rather than concurrent decision support). Measurements in the care of real patients could be clinician-reported (eg, “found an answer”) or objectively assessed and could reflect concurrent support or sustained impact.

We defined *electronic knowledge resource* as quoted in the Introduction, which was adapted from the definition proposed by Lobach [12].We defined clinicians as practitioners or students in a health-related field with direct responsibility for patient-related decisions; this included, but was not limited to, physicians, nurse practitioners, physician assistants, certified nurse anesthetists, pharmacists, midwives, dentists, and psychologists. We excluded studies focused solely on nurses and allied health professionals. We included studies making a comparison with a separate intervention, including randomized, nonrandomized, and crossover designs, or with baseline performance (ie, single-group, pre-/postintervention studies).

Reviewers (DAC, CAA, and LAM) worked independently and in duplicate to screen each identified study for inclusion, first reviewing the title and abstract and then reviewing the full text if needed; the kappa indicating interrater reliability should be greater than or equal to .70. All disagreements were resolved by consensus.

### Data Abstraction

Two reviewers (DAC and LAM) used a standardized abstraction form to independently extract data from all included studies, resolving all disagreements by consensus. We extracted information about the participants, topic, resources used, outcomes, and potential conflicts of interest. We appraised study quality using the Newcastle-Ottawa Scale as modified for education [18,19], which evaluates sample selection and comparability, blinding of assessment, and attrition. We converted all quantitative results, including odds ratios (ORs) [20], to standardized mean differences (SMDs).

### Data Synthesis

We conducted a meta-analysis to pool SMDs whenever three or more studies shared conceptually aligned, between-intervention contrasts [20]. In accordance with our study protocol, we used random-effects meta-analysis because we anticipated pooling across different resources, with likely different effects. We planned to weight studies by the number of users, but most studies reporting clinical outcomes reported only the number of patients or hospitals. Thus, we weighted analyses of knowledge and skills outcomes by the number of users, and we weighted analyses of clinician behaviors and patient effects by the number of patients, with exceptions as noted in the text. We conducted sensitivity analyses limited to randomized trials, recent publications (ie, after 2007), and studies of physicians in practice or postgraduate trainees. We planned to check for publication bias using funnel plots, but the small number of studies precluded meaningful analysis. We estimated heterogeneity using $I^2$.

For studies that did not permit meta-analysis, we synthesized results using narrative methods, taking into account key differences in study design, study quality, intervention, and context.

## Results

We identified 10,811 potentially relevant studies: 10,799 studies in our literature search and 12 from our examination of previous reviews. From these, we included 25 comparative studies evaluating the impact of electronic knowledge resources (see Figure 1) [21-45].

### Study Characteristics

Table 1 summarizes study characteristics and Table 2 provides detailed information about each study. Out of 25 studies, 20 (80%) investigated electronic knowledge resources in the context of patient care, while 5 (20%) took place in laboratory or test settings. Nearly all studies (22/25, 88%) included physicians in practice or in training. Other studies included nurse practitioners or mixed user groups. Common topics included general medicine (15/25, 60%), surgery (5/25, 20%), and pediatrics (5/25, 20%). All studies were published between 2003 and 2016 and were in English.

**Figure 1.** Trial flowchart.

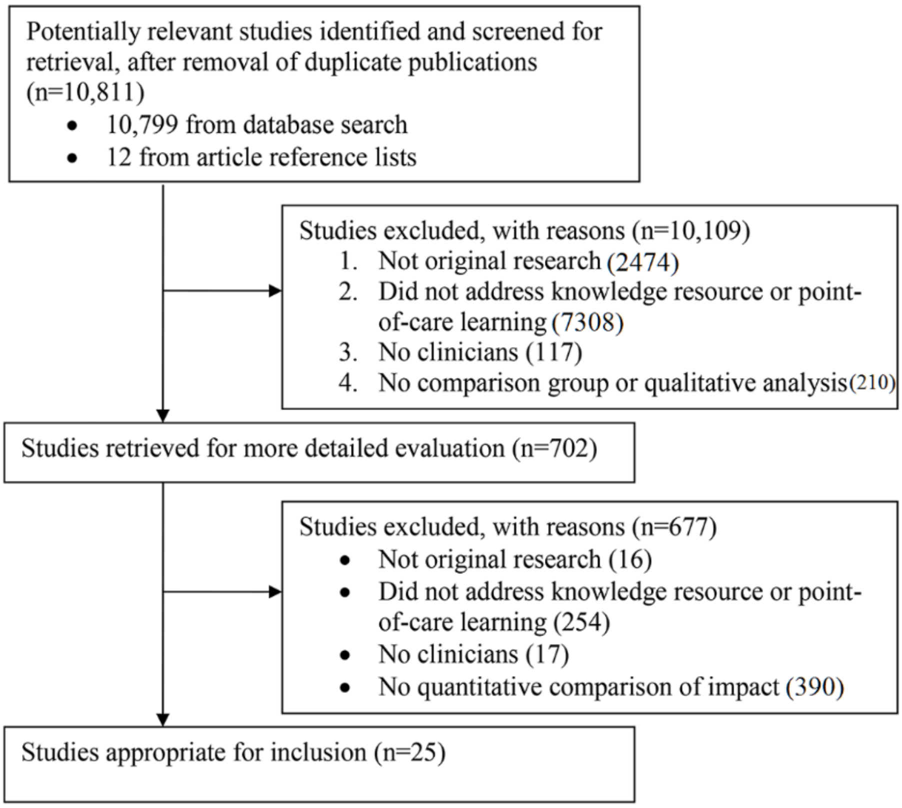


**Table 1.** Summary of key study characteristics and quality.

| Study characteristic | Studies (N=25), n (%) |
|---|---|
| **Participants**[a] | |
| Practicing physicians | 12 (48) |
| Physicians in postgraduate training | 12 (48) |
| Medical students | 4 (16) |
| Nurse practitioners | 3 (12) |
| Mix of user groups | 7 (28) |
| **Clinical topics**[a] | |
| General medicine | 15 (60) |
| Pediatrics | 5 (20) |
| Surgery | 5 (20) |
| Emergency medicine | 3 (12) |
| Medical specialties | 3 (12) |
| Anesthesia | 2 (8) |
| Laboratory medicine, pathology, and radiology | 1 (4) |
| Pharmacy | 1 (4) |
| **Patient setting** | |
| Outpatient | 11 (44) |
| Inpatient | 7 (28) |
| Unspecified | 3 (12) |
| **Electronic knowledge resources**[b] | |
| UpToDate | 6 (24) |
| InfoRetriever | 5 (20) |
| Clinical evidence | 2 (8) |
| DynaMed | 2 (8) |
| Epocrates | 2 (8) |
| MD Consult | 2 (8) |
| Micromedex | 2 (8) |
| Trip[c] | 2 (8) |
| Other | 9 (36) |
| **Comparison resources**[b] | |
| MEDLINE | 4 (16) |
| User choice of any nonknowledge resource | 3 (12) |
| Journals | 3 (12) |
| Paper resources | 3 (12) |
| Curated (eg, library subject guides) | 3 (12) |
| Google | 2 (8) |
| Other search engines | 2 (8) |
| **Quality measures** | |
| **Number of groups** | |
| One study group | 5 (20) |
| Crossover design | 9 (36) |

| Study characteristic | Studies (N=25), n (%) |
|---|---|
| Two or more study groups | 11 (44) |
| **Newcastle-Ottawa Scale results** | |
| Score: ≥4 | 8 (32) |
| Representative: yes | 9 (36) |
| Selection of comparison group: same community | 9/12 (75)[d] |
| Comparability of comparison group: high | 5/12 (42)[d] |
| Follow-up: high (>75%) | 16 (64) |
| Blinded outcomes assessment: yes | 9 (36) |
| **Funding** | |
| Potential conflict of interest | 10 (40) |

[a]The number of studies in some subgroups may add up to more than the total number of studies, and percentages may be more than 100%, because several studies included combinations of clinicians and/or study topics.

[b]Selected list of electronic knowledge resources and comparison resources; other resources were studied with lower frequency.

[c]Trip: Turning Research Into Practice.

[d]Percentage of two-group studies.

**Table 2.** Detailed information about each study.

| Author, year | User type | Topic | Knowledge resource | Comparison | Outcomes |
|---|---|---|---|---|---|
| Leung, 2003 [21] | Medical students | General medical and surgery | InfoRetriever | NI[a] and ORes[b] | Attitudes |
| Schwartz, 2003 [22] | Practicing physicians | General medical | Clinical evidence, DynaMed, InfoRetriever, and Trip[c] | ORes | Knowledge and skills |
| D'Alessandro, 2004 [23] | Practicing physicians and residents | Pediatrics | MD Consult and Micromedex | ORes and KR[d] | Knowledge and skills |
| Alper, 2005 [24] | Practicing physicians and nurse practitioners | General medical and pediatrics | DynaMed, InfoRetriever, Medscape, MD Consult, and UpToDate | ORes | Knowledge and skills |
| Grad, 2005 [25] | Residents | General medical | InfoRetriever | NI | Knowledge and skills |
| Grad, 2005 [26] | Residents | General medical | InfoRetriever | KR | Knowledge and skills |
| Greiver, 2005 [27] | Practicing physicians | General medical | Angina software[e] | NI | Behaviors and patient effect |
| Bochicchio, 2006 [28] | Residents | General medical, surgery, anesthesia, and emergency medicine | Johns Hopkins Antibiotics Guide[e] | NI | Knowledge and skills |
| Maviglia, 2006 [29] | Practicing physicians, residents, and nurse practitioners | General medical and medical specialties | Micromedex | KR | Knowledge and skills |
| Ramnarayan, 2006 [30] | Residents | Pediatrics | Isabel | NI | Behaviors |
| Rudkin, 2006 [31] | Practicing physicians and residents | Emergency medicine | Epocrates, Tarascon Pharmacopeia, The 5-Minute Clinical Consult, and qID | ORes | Knowledge and skills |
| Emery, 2007 [32] | Residents and nurse practitioners | General medical | GRAIDS[e,f] | NI | Behaviors and patient effect |
| King, 2007 [33] | Residents and medical students | Anesthesia and pediatrics | Clinical evidence module[e] | NI | Behaviors and patient effect |
| Magrabi, 2007 [34] | Practicing physicians | General medical | MIMS[g] and Quick Clinical | NI | Attitudes |
| Skeate, 2007 [35] | Residents and medical students | Laboratory medicine, pathology, and radiology | Report Support[e] | NI | Knowledge and skills |
| Van Duppen, 2007 [36] | Practicing physicians and residents | General medical | Clinical evidence and Trip | ORes and KR | Knowledge and skills |
| Bonis, 2008 [37] | Mixed users[h] | Mixed topics | UpToDate | NI | Behaviors and patient effect |
| Hoogendam, 2008 [38] | Practicing physicians and residents | General medical | UpToDate | ORes | Knowledge and skills |
| Lyman, 2008 [39] | Practicing physicians | Pharmacy | Epocrates | NI | Behaviors |
| Isaac, 2012 [40] | Mixed users | General medical and surgery | UpToDate | NI | Behaviors and patient effect |
| Reed, 2012 [41] | Practicing physicians | General medical | PIER[i] and UpToDate | NI and KR | Knowledge and skills |
| Kuhn, 2015 [42] | Mixed users | Medical specialties | eAAP[e,j] | NI | Patient effect |
| Chow, 2016 [43] | Mixed users | General medical, medical specialties, pharmacy, surgery, and mixed topics | ARUSC[e,k] | KR | Behaviors |

| Author, year | User type | Topic | Knowledge resource | Comparison | Outcomes |
|---|---|---|---|---|---|
| Luther, 2016 [44] | Practicing physicians | Emergency medicine, pediatrics, and surgery | SCAMP[e,l] | NI | Behaviors, patient effect, and costs |
| Saparova, 2016 [45] | Medical students | Mixed topics | AccessMedicine, UpToDate, and Wikipedia | KR | Knowledge and skills |

[a]NI: Knowledge resource compared versus no intervention.

[b]ORes: Knowledge resource compared versus other resource.

[c]Trip: Turning Research Into Practice.

[d]KR: Comparison between knowledge resources.

[e]Denotes a locally developed resource.

[f]GRAIDS: Genetic Risk Assessment on the Internet with Decision Support.

[g]MIMS: Monthly Index of Medical Specialties.

[h]An undifferentiated mix of users.

[i]PIER: Physicians' Information and Education Resource.

[j]eAAP: Emergency Asthma Action Plan.

[k]ARUSC: Antibiotic Utilization and Surveillance-Control.

[l]SCAMP: Standardized Clinical Assessment and Management Plans.

The electronic knowledge resources most commonly evaluated were UpToDate (6/25, 24%) and InfoRetriever (5/25, 20%). Several studies evaluated more than one resource. The 25 studies reported 29 distinct contrasts. Out of 29, 15 contrasts (52%) compared electronic knowledge resources with no intervention; 7 (24%) compared electronic knowledge resources with resources not meeting our definition of electronic knowledge resources, such as MEDLINE or a paper resource, hereafter collectively labeled *other resources*; and 7 (24%) compared one electronic knowledge resource against another (eg, Micromedex vs SkolarMD or two implementations of the same resource, such as presentation as a desktop vs mobile version). Across the 29 contrasts, we extracted 48 discrete outcomes, reflecting knowledge and skills (24/48, 50%), behaviors in practice (10/48, 21%), patient effects (10/48, 21%), attitudes (3/48, 6%), and costs (1/48, 2%). Selected contrasts and outcomes are reported in Figures 2 and 3; Multimedia Appendix 2 lists all contrasts and outcomes.

**Figure 2.** Comparative usage of electronic knowledge resources versus no intervention. Knowledge outcome analyses are weighted by user, while behavior and patient effects analyses are weighted by patients or hospitals. “a” denotes a locally developed resource; “b” is the number of hospitals, not patients; “c” indicates no comparison group (ie, one-group, pre-/postintervention study). Abx Guide: Johns Hopkins Antibiotic Guide; Ang Soft: angina software; CEM: clinical evidence module; eAAP: Emergency Asthma Action Plan; Epoc: Epocrates; GRAIDS: Genetic Risk Assessment on the Internet with Decision Support; InfoRet: InfoRetriever; MD: practicing physicians; MOC: Maintenance of Certification; MS: medical students; NP: nurse practitioners; ns: not specified; PG: residents; PIER: Physicians’ Information and Education Resource; Rep Sup: Report Support; SCAMP: Standardized Clinical Assessment and Management Plans; UTD: UpToDate.

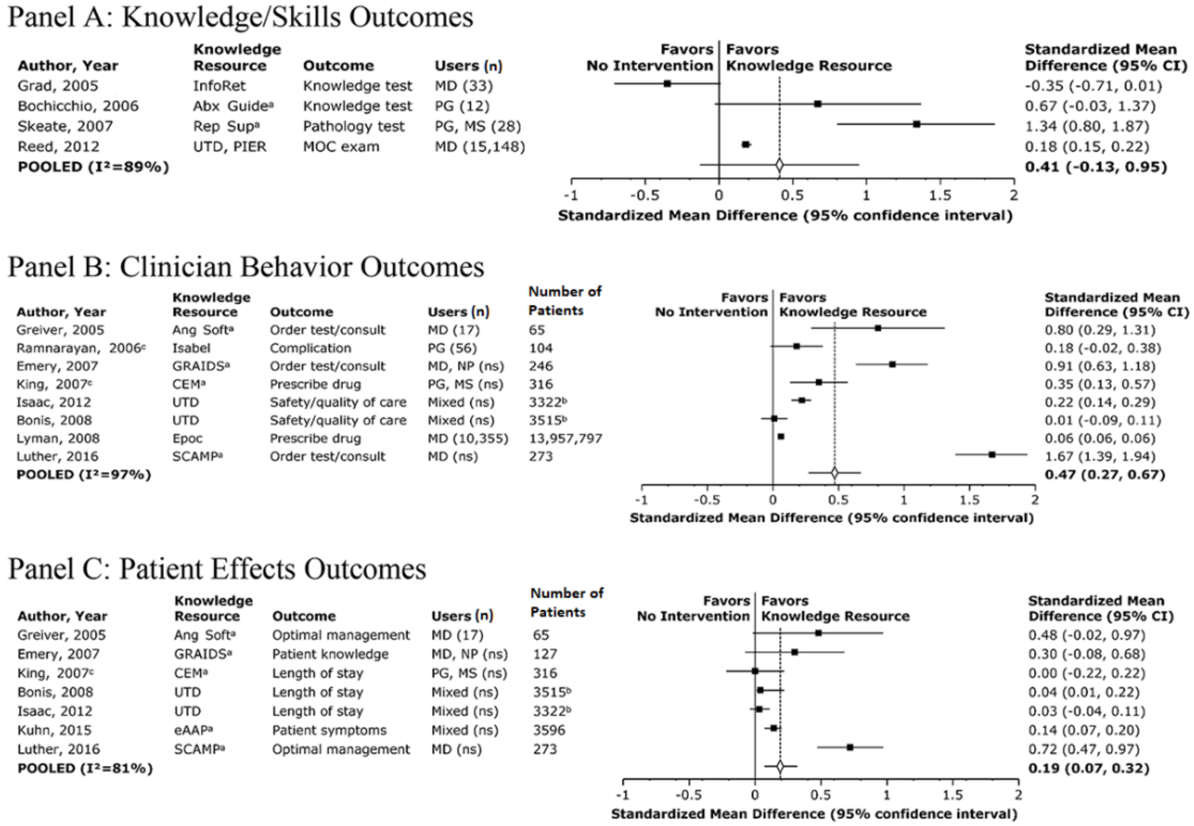
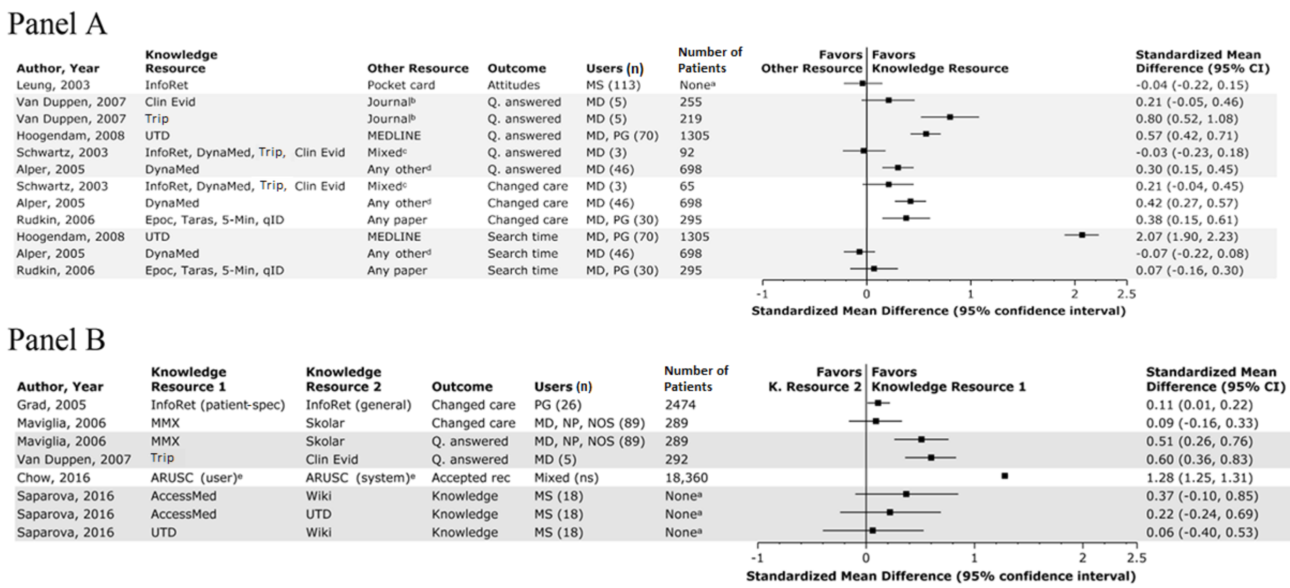



**Figure 3.** Impact of electronic knowledge resources in comparison with other resources (Panel A) and alternate electronic knowledge resources (Panel B). All analyses are weighted by patients except as noted. “a” refers to analysis weighted by users; “b” means the comparison group (ie, study data) is the same for these contrasts; “c” means the comparison type “Mixed” indicates a comparison with both electronic and nonelectronic knowledge resources; “d” means the comparison type “Any other” indicates users could select any resource, except the ones it was being compared against; “e” denotes a locally developed resource. 5-min: 5-Minute Clinical Consult; AccessMed: AccessMedicine; ARUSC: Antibiotic Utilization and Surveillance-Control; Clin Evid: clinical evidence; Epoc: Epocrates; InfoRet: InfoRetriever; K: Knowledge; MD: practicing physicians; MMX: Micromedex; MS: medical students; NOS: not otherwise specified; NP: nurse practitioners; ns: not specified; PG: residents; Q: question; rec: recommendation; spec: specific; Taras: Tarascon Pharmacopeia; Trip: Turning Research Into Practice; UTD: UpToDate; Wiki: Wikipedia.



### Study Quality

When reported, the number of enrolled users ranged from 3 to 15,148; 7 studies out of 25 (28%) did not report the number of users, and 4 (16%) did not report user demographics. A total of 11 (44%) of the 25 studies included two or more groups, of which 5 (45%) were randomized. Assessors were blinded to the study intervention in 9 (36%) of the 25 studies. The mean Newcastle-Ottawa Scale quality score (maximum 6 points) was 2.3 (SD 1.6). In 15 out of 25 studies (60%), outcomes were determined objectively (eg, based on patient records, computer logs, or test scores), including all studies that reported patient outcomes. The other 10 studies (40%) reported only clinician-reported measures (eg, "I found an answer"). Only 9 of 25 studies (36%) enrolled users that were considered representative of the larger community of potential participants. A total of 11 studies (44%) had a separate comparison group; among these, 9 (82%) drew the comparison group from the same community and 5 (45%) were randomized. A total of 16 out of 25 studies (64%) reported high participant follow-up. A total of 10 studies (40%) reported potential financial conflicts of interest (eg, industry grant, discounted or free product pricing, involvement of resource creators, or employment by industry). A total of 6 studies (24%) did not report funding sources (see Table 3).

### Synthesis: Comparisons With No Intervention

A total of 15 studies out of 25 (60%) compared one or more electronic knowledge resources with no intervention, including comparisons of usual practice without versus with access to the resource, reporting a total of 22 outcomes [21,25-28,30,32-35,37,39-42,44]. Of these 15 studies, 9 (60%) reported potential conflicts of interest.

Out of these 15 studies, 4 (27%) reported knowledge or skill outcomes, evaluating InfoRetriever, UpToDate, American College of Physicians (ACP) Physicians' Information and Education Resource (PIER), and three local resources, alone or in varying combinations. The pooled SMD was 0.41 (95% CI –0.13 to 0.95; $P$=.14; see Figure 2, Panel A). Inconsistency was high, with individual SMDs ranging from –0.35 to 1.34 and an $I^2$ of 89%. None of these studies were randomized and only 1 out of the 4 (25%) was published since 2007. Limiting this analysis to the 3 studies out of 4 (75%) without a potential conflict of interest yielded an SMD of 0.35 (95% CI –0.29 to 0.99; $P$=.29). Limiting the analysis to the 3 studies (75%) enrolling physicians in practice or postgraduate trainees revealed an SMD of 0.10 (95% CI –0.34 to 0.54; $P$=.65). Out of the 4 studies, 2 (50%) explored attitudes about information seeking and evidence-based medicine, with results showing improved, neutral, and worsened attitudes, depending on the attitude statement, after use of knowledge resources [21,34].

**Table 3.** Quality appraisal of included studies.

| Author, year | Users, n | Study design | Newcastle-Ottawa Scale score[a] | Representativeness[b] | Comparable cohorts[c] | Follow-up[d] | Objective outcomes[e] | Blinded[f] | COI[g] |
|---|---|---|---|---|---|---|---|---|---|
| Leung, 2003 [21] | 113 | 1 group, crossover | 1 | Yes | N/A[h] | Low | Yes | No | Yes |
| Schwartz, 2003 [22] | 3 | 1 group, crossover | 1 | No | N/A | High | No | No | No |
| D'Alessandro, 2004 [23] | 52 | 1 group, crossover | 1 | No | N/A | High | No | No | No |
| Alper, 2005 [24] | 82 | 1 group, crossover | 1 | No | N/A | High | No | No | Yes |
| Grad, 2005 [25] | 37 | ≥2 groups | 4 | Yes | Similar | High | Yes | No | No |
| Grad, 2005 [26] | 26 | 1 group, crossover | 0 | No | N/A | Low | No | No | No |
| Greiver, 2005 [27] | 18 | ≥2 groups, RCT[i] | 4 | No | Similar | High | Yes | No | No |
| Bochicchio, 2006 [28] | 12 | ≥2 groups, RCT | 4 | No | Similar | High | Yes | Yes | Yes |
| Maviglia, 2006 [29] | 279 | ≥2 groups, RCT | 3 | No | Similar | Low | No | No | Yes |
| Ramnarayan, 2006 [30] | 80 | 1 group | 2 | Yes | N/A | Low | Yes | Yes | Yes |
| Rudkin, 2006 [31] | 30 | 1 group, crossover | 1 | No | N/A | High | No | No | No |
| Emery, 2007 [32] | Not specified | ≥2 groups, RCT | 4 | Yes | Similar | Low | Yes | Yes | Yes |
| King, 2007 [33] | Not specified | 1 group | 0 | No | N/A | Low | Yes | No | No |
| Magrabi, 2007 [34] | 227 | 1 group | 0 | No | N/A | Low | No | No | Yes |
| Skeate, 2007 [35] | 30 | ≥2 groups, RCT | 3 | No | Similar | High | Yes | No | No |
| Van Duppen, 2007 [36] | 5 | 1 group, crossover | 1 | No | N/A | High | No | No | No |
| Bonis, 2008 [37] | Not specified | ≥2 groups | 4 | No | N/A | High | Yes | Yes | Yes |
| Hoogendam, 2008 [38] | 70 | 1 group, crossover | 1 | No | N/A | High | No | No | No |
| Lyman, 2008 [39] | 10,355 | ≥2 groups | 4 | Yes | Similar | High | Yes | Yes | Yes |
| Isaac, 2012 [40] | Not specified | ≥2 groups | 4 | No | N/A | High | Yes | Yes | Yes |
| Reed, 2012 [41] | 15,148 | ≥2 groups | 6 | Yes | Similar | High | Yes | Yes | No |
| Kuhn, 2015 [42] | Not specified | 1 group | 1 | Yes | N/A | Low | Yes | No | No |
| Chow, 2016 [43] | Not specified | 1 group, crossover | 2 | Yes | N/A | Low | No | Yes | No |
| Luther, 2016 [44] | Not specified | 1 group | 2 | Yes | N/A | High | Yes | No | No |
| Saparova, 2016 [45] | 18 | ≥2 groups | 3 | No | Similar | High | Yes | Yes | No |

[a]The score for this scale can be a maximum of 6 points.

[b]"Yes" indicates that the study is truly representative of the average clinician in the community, while "No" indicates that it is not.

[c]"Similar" indicates that the comparison group was drawn from the same community.

[d]"High" indicates that participant follow-up was ≥75%; "Low" indicates that follow-up was <75% or unclear.

[e]"Yes" indicates that at least one outcome was determined objectively; "No" indicates outcomes were not determined objectively.

[f]"Yes" indicates blinded outcomes; "No" indicates no blinding.

[g]"No" indicates no conflict of interest (COI) reported or identified by the reviewer team; "Yes" indicates a reported or identified potential COI.

[h]N/A: not applicable (ie, no separate comparison group).

[i]RCT: randomized controlled trial.

Out of the 15 studies, 8 (53%) reported behavior outcomes, such as appropriate therapy recommendations and test orders, evaluating combinations of Epocrates, Isabel, UpToDate, and four local resources [27,30,32,33,39,40,44]. The pooled SMD was 0.47 (95% CI 0.27 to 0.67; *P*<.001; see Figure 2, Panel B). Inconsistency was again high, with individual SMDs ranging from 0.01 to 1.67 and an $I^2$ of 97%. Out of the 8 studies, 2 (25%) were randomized [27,32]. Limiting analyses to the 4 studies (50%) published since 2007 revealed similar results, with an SMD of 0.41 (95% CI 0.10 to 0.71; *P*=.01). Alternately, limiting to the 3 studies (38%) without a potential conflict of interest yielded an SMD of 0.94 (95% CI 0.02 to 1.86; *P*=.05). Lastly, limiting analysis to the 7 studies (88%) that included physicians in practice or postgraduate trainees produced an SMD of 0.49 (95% CI 0.27 to 0.70; *P*<.001).

Out of the 15 studies, 7 (47%) reported patient effects, including complications, length of stay, optimal management, and mortality, evaluating UpToDate and five locally developed resources [27,32,33,37,40,42,44]. Pooling nonmortality outcomes across these 7 studies revealed an SMD of 0.19 (95% CI 0.07 to 0.32; *P*=.003; see Figure 2, Panel C). Inconsistency was again high, with individual SMDs ranging from 0 to 0.72 and an $I^2$ of 81%. Out of these 7 studies, 2 (29%) were randomized [26,31]. Limiting analyses to the 4 studies out of 7 (57%) published since 2007 revealed a similar SMD of 0.20 (95% CI 0.05 to 0.35; *P*=.01). Limiting to the 4 studies out of 7 (57%) without potential conflicts of interest yielded an SMD of 0.31 (95% CI 0.01 to 0.61; *P*=.04). Focusing on the 6 studies out of 7 (86%) that enrolled physicians in practice or postgraduate trainees produced an SMD of 0.22 (95% CI 0.09 to 0.36; *P*=.001). The 2 studies out of 7 (29%) reporting mortality outcomes, both funded by UpToDate, compared hospitals that did versus did not have access to UpToDate. Out of these 2 studies, 1 (50%) found a very small but statistically significant association between the use of UpToDate and lower mortality (absolute risk difference –0.1%; N=3322 hospitals) [40]; the other found no statistically significant association (risk-adjusted z-score 0.18; N=5515 hospitals) [37].

Out of the 15 studies, 1 (7%) objectively evaluated cost reductions associated with implementation of a local resource; this study found a statistically significant 49% reduction in the cost of care (95% CI 0.46 to 0.53) compared with preimplementation [44].

### Synthesis: Comparisons With Other Resources

A total of 7 studies out of 25 (28%) compared electronic knowledge resources with other information resources that were provided instead of the knowledge resource and that did not meet our definition of electronic knowledge resources (see Figure 3, Panel A) [21-24,31,36,38]. Variation in comparisons and outcomes precluded meaningful meta-analysis. Out of the 7 studies, 2 (29%) found mixed results for the use of electronic knowledge resources on personal digital assistants (PDAs) compared with paper resources. In 1 crossover study (50%), residents given a PDA with electronic knowledge resources (eg, Epocrates and Tarascon Pharmacopeia) demonstrated improvements in self-reported patient management (SMD 0.38, 30 users, 295 patients), compared with resource access limited to print materials [30]. The second study, conducted by the creators of InfoRetriever, found essentially no difference in attitudes about evidence-based medicine when comparing use of a PDA preloaded with InfoRetriever versus an evidence-based medicine pocket card (SMD –0.04, 113 users) [21].

Out of the 7 studies, 2 (29%) suggested that clinicians found answers to more questions, and more rapidly, when using electronic knowledge resources than when using journal-based resources. Out of these 2 studies, 1 crossover study (50%) compared general practitioners' use of Turning Research Into Practice (Trip) and clinical evidence with their use of journal articles from the BMJ and found that these electronic knowledge resources were associated with more frequently finding answers (Trip vs the BMJ: SMD 0.80, 5 users, 219 patients; clinical evidence vs the BMJ: SMD 0.21, 5 users, 255 patients) [36]. Another study (1/2, 50%) reported a statistically significant association between the use of UpToDate and answering more questions (SMD 0.57, 70 users, 1305 patients) and finding answers more quickly (SMD 2.07), in comparison with clinicians using PubMed [38].

Out of the 7 studies, 3 (43%) examined electronic knowledge resources in comparison with a user's choice of any other information resources (eg, Google and textbooks) and reported mixed findings. In a randomized crossover study (1/3, 33%) authored by the founder of DynaMed, physicians using DynaMed reported that they found answers more often (SMD 0.30, 46 users, 698 patients) and that answers more often changed patient care (SMD 0.42), although finding answers took slightly, but not statistically significantly, longer (SMD –0.07) [24]. Another study (1/3, 33%) compared the use of InfoRetriever, DynaMed, Trip, and clinical evidence against a user's choice of any other resources; the study found that the use of these electronic knowledge resources was not significantly associated with clinician-reported success in answering questions (SMD –0.03, 3 users, 92 patients) or changes in care (SMD 0.21, 3 users, 65 patients) [22]. In a third study (1/3, 33%), which is not represented in Figure 3, Panel A, because of insufficient extractable data, pediatricians were randomized to use an online pediatrics library or a resource of their choice and found no statistically significant difference in questions answered or changes in care [23].

### Synthesis: Comparisons Between Electronic Knowledge Resources

The high inconsistency noted in the meta-analyses above may suggest substantial differences between knowledge resources in their implementation (eg, training, policies, and technical support to encourage or facilitate use) and design. Studies comparing different electronic knowledge resources, designs, or implementation strategies can help identify best practices. We identified 7 such studies out of 25 (28%; see Figure 3, Panel B) [23,26,29,36,41,43,45].

Out of these 7 studies, 2 (29%) reported associations between different resource implementation strategies of the same knowledge resource and changes in care. In 1 study (50%), clinicians who were allowed to optionally use a local electronic knowledge resource more often followed the resource's suggestion on antibiotic use compared with when they were provided such information without their request (SMD 1.28, 18,360 patients) [43]. The other study compared two subsections of InfoRetriever: one that employed user-entered patient data to provide patient-specific information and recommendations and the other containing general information resources, such as

The 5-Minute Clinical Consult, Cochrane Reviews, Information Patient-Oriented Evidence that Matters (Info-POEMs), and guideline summaries. This crossover study determined that the patient-specific resources were associated with a slight but statistically significant improvement in clinician-reported changes in care (SMD 0.11, 26 users, 2474 patients) [26].

Out of the 7 studies, 4 (57%) focused on head-to-head comparisons of different electronic knowledge resources. Out of these 4 studies, 1 crossover study (25%) found no statistically significant difference on a knowledge test for 18 medical students who had used Wikipedia, AccessMedicine, or UpToDate (UpToDate vs Wikipedia SMD 0.06; AccessMedicine vs UpToDate SMD 0.22; AccessMedicine vs Wikipedia SMD 0.37) [45]. Another randomized study (1/4, 25%) found no statistically significant differences between Micromedex and SkolarMD in the frequency of answering questions (SMD 0.51, 89 users, 289 patients) or clinician-reported changes in patient care (SMD 0.09) [29]. A randomized crossover study (1/4, 25%) found that clinicians could answer questions more often when using Trip than when using clinical evidence (SMD 0.60, 5 users, 292 patients) [36]. Finally, 1 study (25%) found a statistically significant difference in maintenance of certification exam scores between physicians using two electronic knowledge resources over an extended period; however, due to deliberately blinded reporting, it is not possible to know which resource (ie, PIER or UpToDate) was superior [41]. The effects of these resources in comparison with no intervention were reported earlier in this review.

## Discussion

### Principal Findings

We identified 25 studies that investigated the impact of electronic knowledge resources on patient and clinician outcomes and found results that are mixed and at times contradictory. Nevertheless, we found statistically significant associations between the use of electronic knowledge resources and improved clinician behaviors and patient effects. When compared with other resources, use of electronic knowledge resources was associated with increased success in answering clinical questions, with variable impact on speed. Comparisons of different implementation strategies of the same electronic knowledge resource suggest benefits from allowing clinicians to choose to access the resource, versus automated display of resource information, and from integrating patient-specific information. A total of 4 studies compared different commercial electronic knowledge resources, with variable results.

### Comparison With Other Reviews and Meta-Analyses

Clinicians frequently face clinical questions [1,46], which they are taught and expected to answer using some form of knowledge resources. Previous reviews have focused on interventions to promote knowledge resource adoption [14] or addressed knowledge resources as only one of many information technology tools [12,47]. This review expands upon prior work by focusing specifically on electronic knowledge resources and quantitatively estimating their impact on clinical outcomes and point-of-care learning. Our finding of limited evidence regarding different approaches to electronic knowledge resource implementation strategies parallels the paucity of evidence found in a previous review of health information technology [48].

### Limitations

As with all systematic reviews, our findings are constrained by the quality and quantity of published evidence. For example, only 6 studies reported patient effects and 5 were randomized. Inconsistency was high in all analyses. Additionally, lack of conceptual alignment precluded meta-analysis for comparisons of electronic knowledge resources with other resources or with different implementation strategies. Several studies allowed users access to multiple resources simultaneously, making interpretation difficult. Vague and incomplete reporting limited our ability to extract key information on study design, outcomes, contextual details (eg, setting and disease acuity), and resource design and implementation (eg, how participants accessed the resource, password requirements, or optimization for use on a mobile device) for several studies. A total of 10 studies presented potential conflicts of interest, which could bias results. However, sensitivity analyses limited to recent studies and studies without conflicts of interest generally yielded similar results. The small number of studies precluded meaningful evaluation of publication bias. We did not attempt to distinguish resources based on the developer's intended purpose (eg, education, decision support, or information) but instead focused on the resource's function and application (ie, decision making for a specific patient). Several studies are over a decade old, which limits their relevance to current resource implementations. We conducted our literature search in 2017, and studies published since that date were thus omitted from our analyses. This review has several strengths, including a comprehensive search of multiple databases by information professionals, duplicate review at all stages of screening and data extraction, and broad inclusion criteria encompassing a range of health professionals and topics.

### Implications

In this meta-analysis, use of electronic knowledge resources appeared to improve patient care and their continued use in clinical practice appears to be warranted. More specifically, these resources provide answers to clinician-initiated questions “just in time,” thus preserving clinicians' autonomy and workflow. This functionality contrasts with that of interruptive clinical decision-support systems, such as reminders and alerts, that have been associated with workflow disruption, alert fatigue, inappropriate recommendations, and provider dissatisfaction [49-51]. Use of electronic knowledge resources is also associated with enhanced durable learning (ie, improved performance on knowledge tests conducted without concurrent resource use). Clinicians may benefit from increased and more strategic use of electronic knowledge resources at various stages in training. Knowledge resources may be particularly important for practicing clinicians as part of their lifelong point-of-care learning activities [52]. The optimal promotion of durable learning may require resource features, such as spaced repetition and quizzing [53,54], that differ from those required for concurrent decision support (ie, maximal efficiency). Electronic knowledge resources offer flexibility allowing such features to be built in yet activated only for relevant learners and contexts.

The impact of electronic knowledge resources, while generally favorable, varied widely across studies. Such differences likely arise from specifics of the topic, clinical context, clinician specialty, and clinician stage of training, in addition to the knowledge resource itself. It seems unlikely that any one resource will optimally address all information needs; rather, health care organizations will likely need to make multiple electronic knowledge resources available and effectively integrate electronic knowledge resources into clinician workflows. Suboptimal integration results in suboptimal outcomes, as was seen in one study in this review [43]. Information tools, such as easily accessible online portals (eg, infobuttons [17]), might further help clinicians select resources appropriate for their specific questions and contexts.

Our review highlights several areas for improvement in the quality of research methods and reporting. For example, several studies failed to report the number of participants or participant demographics. Many studies did not use comparison groups, reported limited participant follow-up, or enrolled participants that were not considered representative of the larger community of potential participants. Also, 10 studies presented potential conflicts of interest (eg, funding from a resource vendor). Finally, the majority of studies lacked details on the design and implementation of the resources under investigation, and information was rarely reported regarding the cost—monetary and nonmonetary—of implementation, use, and maintenance. When planning future studies, researchers should consider and seek to mitigate these and other limitations.

Many uncertainties remain regarding optimal design, implementation strategies, and use of electronic knowledge resources. Unfortunately, studies comparing different knowledge resources or making comparisons with no intervention have largely failed to produce generalizable insights in this regard. Additional research is needed to clarify what works, in what context (ie, question type, topic, and clinical setting), and for what outcome. We believe that head-to-head studies of different resources, or different implementation strategies of a given resource, can provide such evidence; however, such studies must be guided by conceptual models and theories (eg, models and theories of information science and translational informatics). Noncomparative studies examining fundamental questions about information seeking, human factors, and user experience will also be useful. Outcomes of costs, both monetary and nonmonetary, will complement outcomes of effectiveness in supporting evidence-based decisions. Attention to these issues will permit more effective design, implementation strategies, and integration into the clinical workflow, which in turn will optimize electronic knowledge resources' benefits to patient care.

### Conclusions

Use of electronic knowledge resources is associated with a positive impact on clinician behaviors and patient effects. Further research into resource design and implementation strategies is needed.


## Acknowledgments

GDF was funded by a National Library of Medicine grant (grant number: 1R01LM011416). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the US Department of Defense, or the US Government.


## Conflicts of Interest

In 2016, LAM received travel funds to deliver a lecture on evidence-based medicine for employees of Ebsco, the parent company of DynaMed; Ebsco did not have any involvement in the conduct of this study. We are unaware of any other conflicts of interest.

## Multimedia Appendix 1

Supplemental search strategies.

[DOCX File, 23KB - jmir_v21i7e13315_app1.docx ]

## Multimedia Appendix 2

Detailed listing of all contrasts and outcomes by study.

[DOCX File, 22KB - jmir_v21i7e13315_app2.docx ]

## References


1. Del Fiol G, Workman TE, Gorman PN. Clinical questions raised by clinicians at the point of care: A systematic review. JAMA Intern Med 2014 May;174(5):710-718. [doi: 10.1001/jamainternmed.2014.368] [Medline: 24663331]
2. Cook DA, Sorensen KJ, Hersh W, Berger RA, Wilkinson JM. Features of effective medical knowledge resources to support point of care learning: A focus group study. PLoS One 2013;8(11):e80318 [FREE Full text] [doi: 10.1371/journal.pone.0080318] [Medline: 24282535]

3. Cook DA, Sorensen KJ, Wilkinson JM, Berger RA. Barriers and decisions when answering clinical questions at the point of care: A grounded theory study. JAMA Intern Med 2013 Nov 25;173(21):1962-1969. [doi: 10.1001/jamainternmed.2013.10103] [Medline: 23979118]
4. Straus SE, Glasziou P, Richardson WS, Haynes RB. Evidence-Based Medicine: How to Practice and Teach EBM. 5th edition. New York, NY: Elsevier; 2019.
5. Aakre CA, Pencille LJ, Sorensen KJ, Shellum JL, Del Fiol G, Maggio LA, et al. Electronic knowledge resources and point-of-care learning: A scoping review. Acad Med 2018 Nov;93(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 57th Annual Research in Medical Education Sessions):S60-S67. [doi: 10.1097/ACM.0000000000002375] [Medline: 30365431]
6. Ketchum AM, Saleh AA, Jeong K. Type of evidence behind point-of-care clinical information products: A bibliometric analysis. J Med Internet Res 2011 Feb 18;13(1):e21 [FREE Full text] [doi: 10.2196/jmir.1539] [Medline: 21335319]
7. Maggio LA, Cate OT, Moorhead LL, van Stiphout F, Kramer BM, Ter Braak E, et al. Characterizing physicians' information needs at the point of care. Perspect Med Educ 2014 Nov;3(5):332-342 [FREE Full text] [doi: 10.1007/s40037-014-0118-z] [Medline: 24865885]
8. Haynes RB, Holland J, Cotoi C, McKinlay RJ, Wilczynski NL, Walters LA, et al. McMaster PLUS: A cluster randomized clinical trial of an intervention to accelerate clinical use of evidence-based information from digital libraries. J Am Med Inform Assoc 2006;13(6):593-600 [FREE Full text] [doi: 10.1197/jamia.M2158] [Medline: 16929034]
9. Murray H. More than 2 billion pairs of eyeballs: Why aren't you sharing medical knowledge on Wikipedia? BMJ Evid Based Med 2019 Jun;24(3):90-91. [doi: 10.1136/bmjebm-2018-111040] [Medline: 30108058]
10. Barreau D, Bouton C, Renard V, Fournier J. Health sciences libraries' subscriptions to journals: Expectations of general practice departments and collection-based analysis. J Med Libr Assoc 2018 Apr;106(2):235-243 [FREE Full text] [doi: 10.5195/jmla.2018.282] [Medline: 29632446]
11. Pluye P, Grad RM, Dunikowski LG, Stephenson R. Impact of clinical information-retrieval technology on physicians: A literature review of quantitative, qualitative and mixed methods studies. Int J Med Inform 2005 Sep;74(9):745-768. [doi: 10.1016/j.ijmedinf.2005.05.004] [Medline: 15996515]
12. Lobach D, Sanders GD, Bright TJ, Wong A, Dhurjati R, Bristow E, et al. Enabling health care decision making through clinical decision support and knowledge management. Evid Rep Technol Assess (Full Rep) 2012 Apr(203):1-784. [Medline: 23126650]
13. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: A systematic review. Ann Intern Med 2012 Jul 03;157(1):29-43. [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
14. Gagnon M, Légaré F, Labrecque M, Frémont P, Pluye P, Gagnon J, et al. Interventions for promoting information and communication technologies adoption in healthcare professionals. Cochrane Database Syst Rev 2009 Jan 21(1):CD006093 [FREE Full text] [doi: 10.1002/14651858.CD006093.pub2] [Medline: 19160265]
15. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, et al. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: A literature review. Health Info Libr J 2013 Sep;30(3):178-190 [FREE Full text] [doi: 10.1111/hir.12036] [Medline: 23981019]
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
17. Cook DA, Teixeira MT, Heale BS, Cimino JJ, Del Fiol G. Context-sensitive decision support (infobuttons) in electronic health records: A systematic review. J Am Med Inform Assoc 2017 Mar 01;24(2):460-468 [FREE Full text] [doi: 10.1093/jamia/ocw104] [Medline: 27497794]
18. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Ottawa Hospital Research Institute. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses URL: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [accessed 2019-05-09] [WebCite Cache ID 78FQtxEz5]
19. Cook DA, Reed DA. Appraising the quality of medical education research methods: The Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. Acad Med 2015 Aug;90(8):1067-1076. [doi: 10.1097/ACM.0000000000000786] [Medline: 26107881]
20. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. Stat Med 2001 Mar 30;20(6):825-840. [doi: 10.1002/sim.650] [Medline: 11252006]
21. Leung GM, Johnston JM, Tin KY, Wong IO, Ho LM, Lam WW, et al. Randomised controlled trial of clinical decision support tools to improve learning of evidence based medicine in medical students. BMJ 2003 Nov 08;327(7423):1090 [FREE Full text] [doi: 10.1136/bmj.327.7423.1090] [Medline: 14604933]
22. Schwartz K, Northrup J, Israel N, Crowell K, Lauder N, Neale AV. Use of online evidence-based resources at the point of care. Fam Med 2003 Apr;35(4):251-256. [Medline: 12729308]
23. D'Alessandro DM, Kreiter CD, Peterson MW. An evaluation of information-seeking behaviors of general pediatricians. Pediatrics 2004 Jan;113(1 Pt 1):64-69. [doi: 10.1542/peds.113.1.64] [Medline: 14702450]

24. Alper BS, White DS, Ge B. Physicians answer more clinical questions and change clinical decisions more often with synthesized evidence: A randomized trial in primary care. Ann Fam Med 2005;3(6):507-513 [FREE Full text] [doi: 10.1370/afm.370] [Medline: 16338914]
25. Grad RM, Meng Y, Bartlett G, Dawes M, Pluye P, Boillat M, et al. Effect of a PDA-assisted evidence-based medicine course on knowledge of common clinical problems. Fam Med 2005;37(10):734-740. [Medline: 16273453]
26. Grad RM, Pluye P, Meng Y, Segal B, Tamblyn R. Assessing the impact of clinical information-retrieval technology in a family practice residency. J Eval Clin Pract 2005 Dec;11(6):576-586. [doi: 10.1111/j.1365-2753.2005.00594.x] [Medline: 16364111]
27. Greiver M, Drummond N, White D, Weshler J, Moineddin R, North Toronto Primary Care Research Network (Nortren). Angina on the Palm: Randomized controlled pilot trial of Palm PDA software for referrals for cardiac testing. Can Fam Physician 2005 Mar;51:382-383 [FREE Full text] [Medline: 16926933]
28. 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]
29. Maviglia SM, Yoon CS, Bates DW, Kuperman G. KnowledgeLink: Impact of context-sensitive information retrieval on clinicians' information needs. J Am Med Inform Assoc 2006;13(1):67-73 [FREE Full text] [doi: 10.1197/jamia.M1861] [Medline: 16221942]
30. Ramnarayan P, Winrow A, Coren M, Nanduri V, Buchdahl R, Jacobs B, et al. Diagnostic omission errors in acute paediatric practice: Impact of a reminder system on decision-making. BMC Med Inform Decis Mak 2006 Nov 06;6:37 [FREE Full text] [doi: 10.1186/1472-6947-6-37] [Medline: 17087835]
31. Rudkin SE, Langdorf MI, Macias D, Oman JA, Kazzi AA. Personal digital assistants change management more often than paper texts and foster patient confidence. Eur J Emerg Med 2006 Apr;13(2):92-96. [doi: 10.1097/01.mej.0000192049.04729.0c] [Medline: 16525237]
32. Emery J, Morris H, Goodchild R, Fanshawe T, Prevost AT, Bobrow M, et al. The GRAIDS Trial: A cluster randomised controlled trial of computer decision support for the management of familial cancer risk in primary care. Br J Cancer 2007 Aug 20;97(4):486-493 [FREE Full text] [doi: 10.1038/sj.bjc.6603897] [Medline: 17700548]
33. King WJ, Le Saux N, Sampson M, Gaboury I, Norris M, Moher D. Effect of point of care information on inpatient management of bronchiolitis. BMC Pediatr 2007 Jan 24;7:4 [FREE Full text] [doi: 10.1186/1471-2431-7-4] [Medline: 17250764]
34. Magrabi F, Westbrook JI, Coiera EW. What factors are associated with the integration of evidence retrieval technology into routine general practice settings? Int J Med Inform 2007 Oct;76(10):701-709. [doi: 10.1016/j.ijmedinf.2006.06.009] [Medline: 16893677]
35. Skeate RC, Wahi MM, Jessurun J, Connelly DP. Personal digital assistant-enabled report content knowledgebase results in more complete pathology reports and enhances resident learning. Hum Pathol 2007 Dec;38(12):1727-1735. [doi: 10.1016/j.humpath.2007.05.019] [Medline: 17949779]
36. Van Duppen D, Aertgeerts B, Hannes K, Neirinckx J, Seuntjens L, Goossens F, et al. Online on-the-spot searching increases use of evidence during consultations in family practice. Patient Educ Couns 2007 Sep;68(1):61-65. [doi: 10.1016/j.pec.2007.04.008] [Medline: 17540531]
37. Bonis PA, Pickens GT, Rind DM, Foster DA. Association of a clinical knowledge support system with improved patient safety, reduced complications and shorter length of stay among Medicare beneficiaries in acute care hospitals in the United States. Int J Med Inform 2008 Nov;77(11):745-753. [doi: 10.1016/j.ijmedinf.2008.04.002] [Medline: 18565788]
38. Hoogendam A, Stalenhoef AF, Robbé PF, Overbeke AJ. Answers to questions posed during daily patient care are more likely to be answered by UpToDate than PubMed. J Med Internet Res 2008 Oct 03;10(4):e29 [FREE Full text] [doi: 10.2196/jmir.1012] [Medline: 18926978]
39. Lyman JA, Conaway M, Lowenhar S. Formulary access using a PDA-based drug reference tool: Does it affect prescribing behavior? AMIA Annu Symp Proc 2008 Nov 06:1034. [Medline: 18998942]
40. Isaac T, Zheng J, Jha A. Use of UpToDate and outcomes in US hospitals. J Hosp Med 2012 Feb;7(2):85-90. [doi: 10.1002/jhm.944] [Medline: 22095750]
41. Reed DA, West CP, Holmboe ES, Halvorsen AJ, Lipner RS, Jacobs C, et al. Relationship of electronic medical knowledge resource use and practice characteristics with Internal Medicine Maintenance of Certification Examination scores. J Gen Intern Med 2012 Aug;27(8):917-923 [FREE Full text] [doi: 10.1007/s11606-012-2017-5] [Medline: 22374410]
42. Kuhn L, Reeves K, Taylor Y, Tapp H, McWilliams A, Gunter A, et al. Planning for action: The impact of an asthma action plan decision support tool integrated into an electronic health record (EHR) at a large health care system. J Am Board Fam Med 2015;28(3):382-393 [FREE Full text] [doi: 10.3122/jabfm.2015.03.140248] [Medline: 25957371]
43. Chow AL, Ang A, Chow CZ, Ng TM, Teng C, Ling LM, et al. Implementation hurdles of an interactive, integrated, point-of-care computerised decision support system for hospital antibiotic prescription. Int J Antimicrob Agents 2016 Feb;47(2):132-139. [doi: 10.1016/j.ijantimicag.2015.12.006] [Medline: 26774157]

44. Luther G, Miller PE, Mahan ST, Waters PM, Bae DS. Decreasing resource utilization using standardized clinical assessment and management plans (SCAMPs). J Pediatr Orthop 2016 Sep 15;39(4):169-174. [doi: 10.1097/BPO.0000000000000873] [Medline: 27636917]
45. Saparova D, Nolan NS. Evaluating the appropriateness of electronic information resources for learning. J Med Libr Assoc 2016 Jan;104(1):24-32 [FREE Full text] [doi: 10.3163/1536-5050.104.1.004] [Medline: 26807049]
46. Ely JW, Burch RJ, Vinson DC. The information needs of family physicians: Case-specific clinical questions. J Fam Pract 1992 Sep;35(3):265-269. [Medline: 1517722]
47. Bright MA. The National Cancer Institute's Cancer Information Service: A premiere cancer information and education resource for the nation. J Cancer Educ 2007;22(1 Suppl):S2-S7. [doi: 10.1080/08858190701348075] [Medline: 17571997]
48. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: An updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 07;160(1):48-54. [doi: 10.7326/M13-1531] [Medline: 24573664]
49. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation. Health Aff (Millwood) 2011 Dec;30(12):2310-2317. [doi: 10.1377/hlthaff.2010.1111] [Medline: 22147858]
50. Wright A, Ai A, Ash J, Wiesen JF, Hickman TT, Aaron S, et al. Clinical decision support alert malfunctions: Analysis and empirically derived taxonomy. J Am Med Inform Assoc 2018 May 01;25(5):496-506 [FREE Full text] [doi: 10.1093/jamia/ocx106] [Medline: 29045651]
51. Wright A, Hickman TT, McEvoy D, Aaron S, Ai A, Andersen JM, et al. Analysis of clinical decision support system malfunctions: A case series and survey. J Am Med Inform Assoc 2016 Dec;23(6):1068-1076 [FREE Full text] [doi: 10.1093/jamia/ocw005] [Medline: 27026616]
52. Cook DA, Blachman MJ, Price DW, West CP, Berger RA, Wittich CM. Professional development perceptions and practices among US physicians: A cross-specialty national survey. Acad Med 2017 Dec;92(9):1335-1345. [doi: 10.1097/ACM.0000000000001624] [Medline: 28225460]
53. Larsen DP, Butler AC, Roediger HL. Repeated testing improves long-term retention relative to repeated study: A randomised controlled trial. Med Educ 2009 Dec;43(12):1174-1181. [doi: 10.1111/j.1365-2923.2009.03518.x] [Medline: 19930508]
54. Larsen DP, Butler AC, Roediger HL. Comparative effects of test-enhanced learning and self-explanation on long-term retention. Med Educ 2013 Jul;47(7):674-682. [doi: 10.1111/medu.12141] [Medline: 23746156]

## Abbreviations

**ACP:** American College of Physicians
**ARUSC:** Antibiotic Utilization and Surveillance-Control
**COI:** conflict of interest
**eAAP:** Emergency Asthma Action Plan
**GRAIDS:** Genetic Risk Assessment on the Internet with Decision Support
**Info-POEMs:** Information Patient-Oriented Evidence that Matters
**KR:** comparison between knowledge resources
**MIMS:** Monthly Index of Medical Specialties
**NI:** knowledge resource compared versus no intervention
**OR:** odds ratio
**ORes:** knowledge resource compared versus other resource
**PDA:** personal digital assistant
**PIER:** Physicians’ Information and Education Resource
**PLUS:** Premium LiteratUre Service
**RCT:** randomized controlled trial
**SCAMP:** Standardized Clinical Assessment and Management Plans
**SMD:** standardized mean difference
**Trip:** Turning Research Into Practice

*Edited by G Eysenbach; submitted 05.01.19; peer-reviewed by C White-Williams, H Hah, S Peters, O Ntsweng; comments to author 27.04.19; revised version received 12.05.19; accepted 18.06.19; published 25.07.19.*

*Please cite as:*
*Maggio LA, Aakre CA, Del Fiol G, Shellum J, Cook DA*
*Impact of Clinicians' Use of Electronic Knowledge Resources on Clinical and Learning Outcomes: Systematic Review and Meta-Analysis*
*J Med Internet Res 2019;21(7):e13315*
*URL: http://www.jmir.org/2019/7/e13315/*
*doi:10.2196/13315*
*PMID:31359865*

Original Paper

# A Real-Time Early Warning System for Monitoring Inpatient Mortality Risk: Prospective Study Using Electronic Medical Record Data

Chengyin Ye[1*], PhD; Oliver Wang[2*], BA; Modi Liu[2], BS; Le Zheng[3,4], PhD; Minjie Xia[2], BS; Shiying Hao[3,4], PhD; Bo Jin[2], MS; Hua Jin[2], MS; Chunqing Zhu[2], MS; Chao Jung Huang[5], PhD; Peng Gao[6,7], PhD; Gray Ellrodt[8], MD; Denny Brennan[9], MEd, MBA; Frank Stearns[2], MHA; Karl G Sylvester[7], MD; Eric Widen[2], MHA; Doff B McElhinney[3,4], MD; Xuefeng Ling[4,7], PhD

[1]Department of Health Management, Hangzhou Normal University, Hangzhou, China
[2]HBI Solutions Inc, Palo Alto, CA, United States
[3]Department of Cardiothoracic Surgery, Stanford University, Stanford, CA, United States
[4]Clinical and Translational Research Program, Betty Irene Moore Children's Heart Center, Lucile Packard Children’s Hospital, Palo Alto, CA, United States
[5]National Taiwan University-Stanford Joint Program Office of AI in Biotechnology, Ministry of Science and Technology Joint Research Center for Artificial Intelligence Technology and All Vista Healthcare, Taipei, Taiwan
[6]Shandong University of Traditional Chinese Medicine, Shandong, China
[7]Department of Surgery, Stanford University, Stanford, CA, United States
[8]Department of Medicine, Berkshire Medical Center, Pittsfield, MA, United States
[9]Massachusetts Health Data Consortium, Waltham, CA, United States
*these authors contributed equally

**Corresponding Author:**
Xuefeng Ling, PhD
Department of Surgery
Stanford University
S370 Grant Bldg, 300 Pasteur Drive
Stanford, CA, 94305
United States
Phone: 1 6504279198
Email: bxling@stanford.edu

## Abstract

**Background:** The rapid deterioration observed in the condition of some hospitalized patients can be attributed to either disease progression or imperfect triage and level of care assignment after their admission. An early warning system (EWS) to identify patients at high risk of subsequent intrahospital death can be an effective tool for ensuring patient safety and quality of care and reducing avoidable harm and costs.

**Objective:** The aim of this study was to prospectively validate a real-time EWS designed to predict patients at high risk of inpatient mortality during their hospital episodes.

**Methods:** Data were collected from the system-wide electronic medical record (EMR) of two acute Berkshire Health System hospitals, comprising 54,246 inpatient admissions from January 1, 2015, to September 30, 2017, of which 2.30% (1248/54,246) resulted in intrahospital deaths. Multiple machine learning methods (linear and nonlinear) were explored and compared. The tree-based random forest method was selected to develop the predictive application for the intrahospital mortality assessment. After constructing the model, we prospectively validated the algorithms as a real-time inpatient EWS for mortality.

**Results:** The EWS algorithm scored patients’ daily and long-term risk of inpatient mortality probability after admission and stratified them into distinct risk groups. In the prospective validation, the EWS prospectively attained a c-statistic of 0.884, where 99 encounters were captured in the highest risk group, 69% (68/99) of whom died during the episodes. It accurately predicted the possibility of death for the top 13.3% (34/255) of the patients at least 40.8 hours before death. Important clinical utilization features, together with coded diagnoses, vital signs, and laboratory test results were recognized as impactful predictors in the final EWS.

**Conclusions:** In this study, we prospectively demonstrated the capability of the newly-designed EWS to monitor and alert clinicians about patients at high risk of in-hospital death in real time, thereby providing opportunities for timely interventions. This real-time EWS is able to assist clinical decision making and enable more actionable and effective individualized care for patients' better health outcomes in target medical facilities.

***(J Med Internet Res 2019;21(7):e13719)*** doi:10.2196/13719

**KEYWORDS**

inpatients; mortality; risk assessment; electronic health records; machine learning

## Introduction

### Importance of an Early Warning System

The condition of some hospitalized patients rapidly deteriorates because of either disease progression or imperfect triage and level of care assignment after their admission. Evidence from observational studies show that signs of clinical deterioration (eg, abnormal vital signs) hours before a serious clinical event [1,2] are important predictors. Therefore, we hypothesize that an early warning system (EWS) to identify patients at high risk of subsequent intrahospital death can be an effective tool to improve patient safety and quality of care and also reduce avoidable harm and costs. For patients without a do-not-resuscitate (DNR) order, a warning from such an EWS can activate rapid response teams (RRTs) or medical emergency teams to offer more intensive care and enhanced attention to prevent hospital deaths [3,4]. For those patients with a DNR order, the notification can trigger health caregivers to counsel and work with patients' families to initiate the end-of-life care and deathbed farewell [5]. Therefore, an EWS to identify and alert caregivers of truly high-risk patients before their deterioration is recognized as an essential step toward the advancement of individualized medical interventions, the improvement of end-of-life patient care quality, and the reduction of unnecessary in-hospital mortality and associated health resource utilization.

### Current Development of an Early Warning System

During the last decade, an increasing number of hospital systems have started to implement EWSs to monitor all adult patients in acute hospital settings and to identify adverse trends and patient deterioration [3,6]. A variety of EWSs have been developed using patients' postadmission clinical information [2,7-14]. For instance, the widely used VitalPAC Early Warning Score (ViEWS) is calculated from 7 vital sign parameters selected by a thorough literature review and was proven to outperform most other published systems when predicting in-hospital death within 24 hours postobservation [9,10]. However, in addition to using a limited number of parameters empirically selected by experts, we speculate whether such EWSs could be further improved in terms of both sensitivity and specificity by integrating more clinical and nonclinical information (eg, disease diagnoses and social determinant data).

With the rapid growth of hospital adoption of electronic medical record (EMR) systems, other temporal clinical information is becoming available at the point of care and can be used to facilitate the prediction of in-hospital mortality. Several EMR-based risk models have been constructed using surgical record data, laboratory test results, and location transfer information [13-17]. Among these models, clinical utilization factors were rarely considered as potential predictors for inpatient mortality, even though they are valuable features from the aspect of hospital quality measurement.

### Aim of This Study

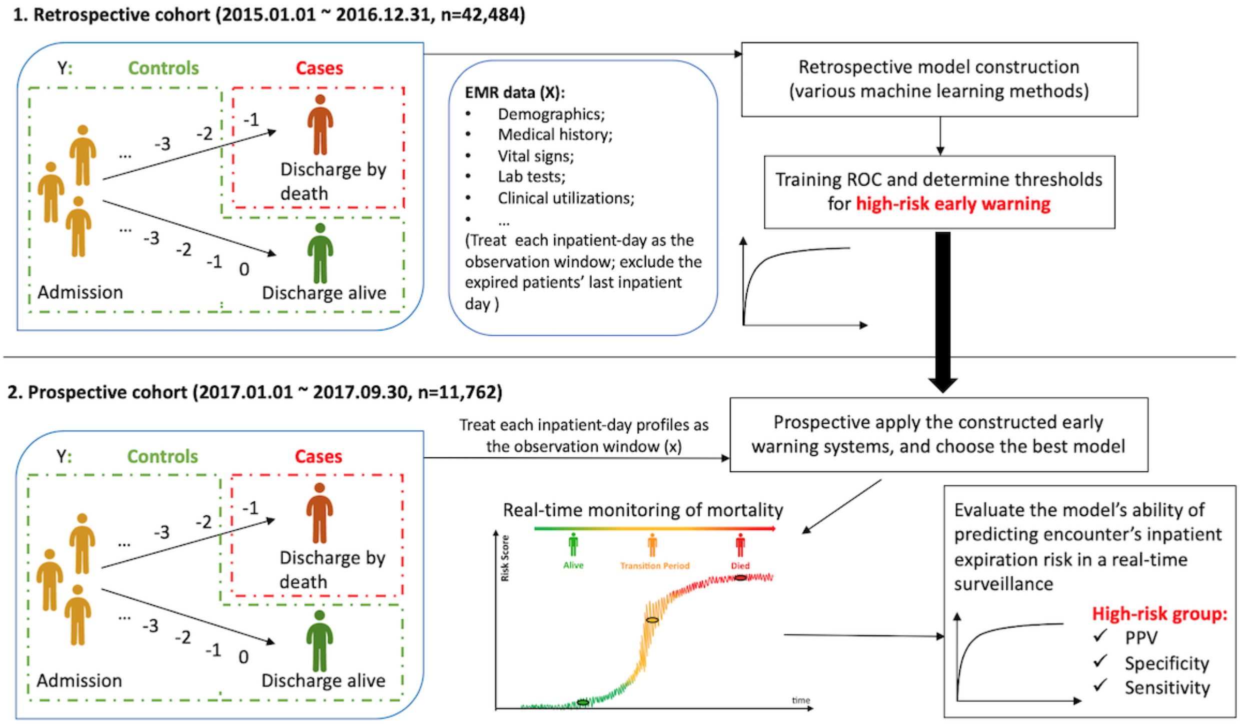
In this study, we aimed to build and prospectively validate an EMR-based inpatient mortality EWS. We expected that by adopting the machine learning algorithms, this EMR-based EWS could capture previously ignored but useful variables and attain an improved discriminative ability with higher levels of sensitivity and specificity. We evaluated its performance and addressed questions such as how early the system can prospectively alarm for a hospitalized mortality event. We also studied the association between impactful predictors (eg, historical clinical utilizations) and inpatient mortality under various circumstances.

## Methods

### Setting and Patient Population

The patient population is defined as the patients admitted to two acute hospitals (ie, Berkshire Medical Center and Fairview Hospital) within the Berkshire Health System (BHS), between January 1, 2015 and September 30, 2017. Patients included in the study were those admitted to a medical unit (including the intensive care unit) from either the emergency department (ED) or outpatient clinics, regardless of whether they had DNR orders. The details are shown in the study design workflow (Figure 1). BHS authorized the use of the deidentified data for this research, and thus all personal privacy information was masked during the process of analysis and publication. This study was also exempted from ethics review by the Stanford University institutional review board (September 25, 2018).

**Figure 1.** Study design. The early warning system model was built on the retrospective cohort (n=42,484) and validated on the prospective cohort (n=11,762). EMR: electronic medical record; PPV: positive predictive value; ROC: receiver operating characteristic.




## Outcome Variables

Following the rationale that patients who died showed signs of clinical deterioration before death, we identified the cases of our study as the 24-hour period immediately before the day of death for those patients who died and classified all other 24-hour periods as the controls (Figure 1). We used the EMR profile collected before the future 24-hour period as the predictors of the following 24-hour period, making the prediction model capable of estimating the risk of death at least 24 to 48 hours before the event.

## Predictor Variables and Feature Selection

In this study, we defined an inpatient day as a time period between 12:00 am and 11:59 pm in an episode. Within an encounter's observation window (ie, each inpatient day), candidate predictor variables were extracted from the hospital EMR system, comprising (1) a set of static historical medical variables and (2) a number of dynamic updated postadmission clinical information. By using the medical data cumulatively collected until a certain inpatient day after admission, the risk model was initially designed to predict a patient's probability of dying in the following inpatient day. Before the machine learning process, we carried out feature selection using both literature review for including impactful feature inclusion and a univariate filtering process for exclusion. As a result, we recruited 680 potential predictors into the subsequent analysis.

## Retrospective Derivation and Prospective Evaluation of the Real-Time Inpatient Mortality Early Warning System

### *Retrospective Model Derivation*

At the derivation stage, the real-time inpatient mortality model was constructed on the EMR data collected within a retrospective 2-year period during January 1, 2015 and December 31, 2016, with a total of 42,484 inpatient encounters (Figure 1). At this stage, multiple existing predictive machine learning algorithms (linear and nonlinear) were explored to construct the prediction model, including the tree-based random forest method [18], XGBoost [19], Boosting [20], Support Vector Machine [21], LASSO [22], and K-nearest neighbors [23]. Following this, the predicted outcomes were calibrated to the positive predictive values (PPVs) on the retrospective cohort. This allowed us to calculate the risk score of mortality for each inpatient day during the in-hospital episode and use the quintiles of these calibrated risk scores to stratify risk groups. The propensity score matching was also introduced to investigate the causal relation between high-weight chronic-based risk factors and the inpatient mortality outcome.

### *Prospective Model Evaluation*

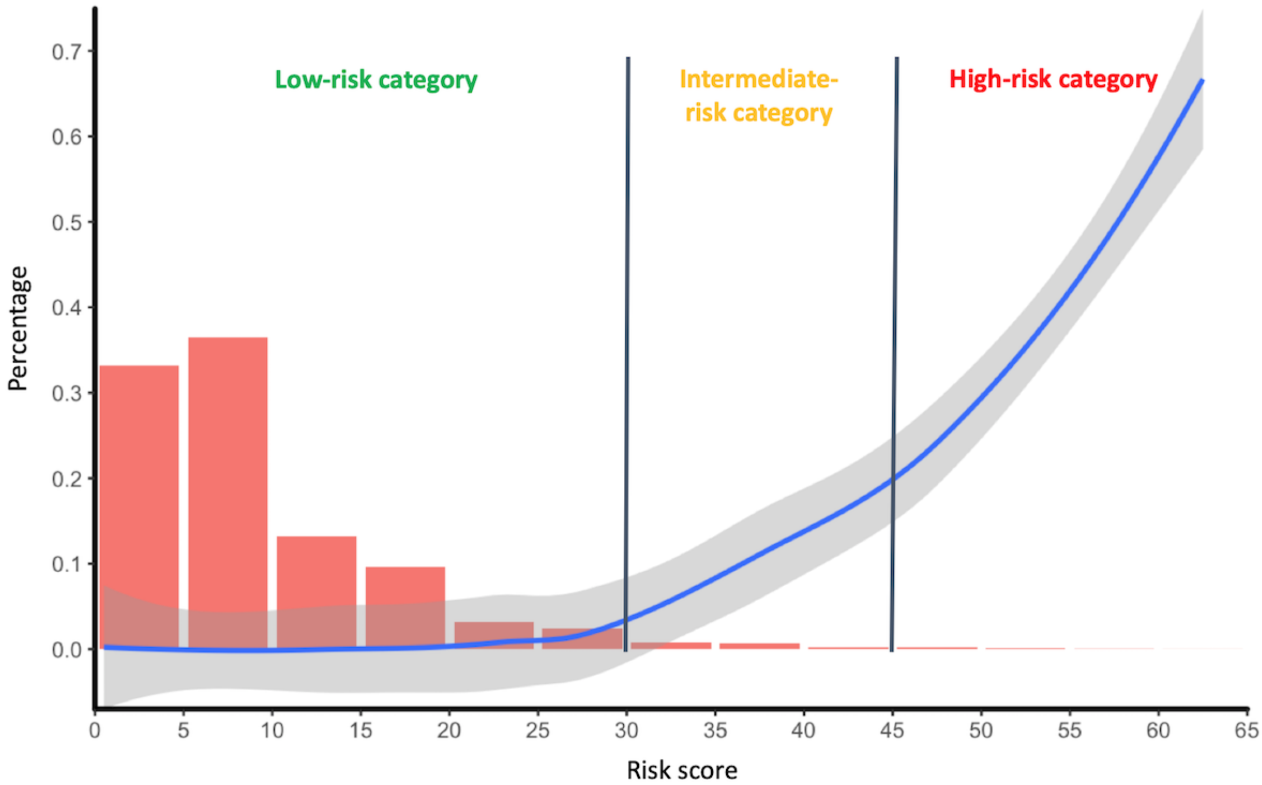
The constructed models were prospectively evaluated on inpatient admissions for the period January 1, 2017 to September 30, 2017. A total of 11,762 hospitalized patients were assigned an EWS score during this period. The discriminatory power of various algorithms was assessed and compared using the receiver operating characteristic (ROC) curve and the prospectively validated c-statistic. According to the prospective results, the model that attained the best performance was chosen as the proposed EWS. Using the final EWS, we also derived the distribution of inpatient days across the spectrum of the calibrated risk scores and evaluated various risk bins for sensitivity, specificity, and PPVs. On the basis of these determined risk categories (*low*, *intermediate* and *high*), we prospectively explored their subsequent mortality rate using the Kaplan-Meier method and compared their hazard ratios (HRs) using Cox regression. We also conducted subgroup analysis to review the model's utility on encounters with specific conditions (eg, DNR orders or high clinical costs in the past).

## Results

### Inpatient Mortality Early Warning System Performance on Inpatient-Day Level

The retrospective and prospective cohorts comprised 42,484 and 11,762 encounters, respectively; 2.34% (993/42,484) and 2.17% (255/11,762) of the patients in these cohorts died during their episode. The demographics and important characteristics of these two cohorts were summarized in Multimedia Appendix 1. After applying the various EWS algorithms to the prospective cohort, we compared their performance as measured by the ROC curve and validated c-statistic. The tree-based random forest algorithm attained the highest predicted c-statistic of 0.884, whereas other machine learning algorithms (linear and nonlinear) attained a predicted c-statistic between 0.511 and 0.867 (Multimedia Appendix 2). Thus, we chose the random forest algorithm–based EWS as the final proposed EWS, where we initially assigned a calibrated risk score to each inpatient day and then stratified these inpatient days into distinct risk groups across the spectrum of risk scores (Figure 2). For a total of 56,588 observed inpatient days, almost 69.66% (39,420/56,588) were located in the *low-risk* percentiles (ie, 0-10), with only 0.09% (35/39,420) of them being cases. Meanwhile, a total of 189 observations fell into the *high-risk* percentiles (ie, ≥45), with 31.2% (59/189) passing away in the subsequent 24 hours (Multimedia Appendix 3).

**Figure 2.** The distribution of inpatient days (the red bar) and positive predictive values (the blue line), coordinated with the inpatient-mortality risk scores on the prospective cohort.




### Performance of the Early Warning System in Predicting Patients' Overall Inpatient Mortality

In terms of long-term in-hospital mortality, the proposed EWS model captured 99 encounters with *high risk* of expiration (ie, risk score ≥45) and recognized 327 encounters as *intermediate-risk* individuals (ie, risk score 30-45) at the prospective validation stage (Multimedia Appendix 4). By further tracking the *high-risk* patients' mortality rate for the subsequent 20 days, we confirmed that the EWS model successfully alerted clinicians to 40% (40/99) of the top risk encounters 24 to 48 hours before their death, notified another 17% (17/99) 48 to 72 hours before their death, and identified the remaining 11% (11/99) 3 to 7 days ahead of their death, making the survival probability drop to 0.24 within 1 week after triggering the alarms (Figure 3). Furthermore, the mortality hazard ratio of the *high-risk* category is as high as 93.65 (95% CI 68.75-127.57) for the subsequent 20-day time period compared with that of the *low-risk* category. In addition, when focusing on the patients who passed away, the results illustrated that the EWS model successfully seized the top 13.3% (34/255) of the population at least 1.7 inpatient days (40.8 hours) before their death (Figure 4). These findings demonstrated that the proposed EWS had powerful discriminative ability to help notify caregivers of inpatient death in the longitudinal scale and assist in clinical decision making.

**Figure 3.** The observed survival curves of the 3 risk categories (encounter-level) stratified by the real-time early warning system in the prospective validation cohort. HR: hazard ratio.

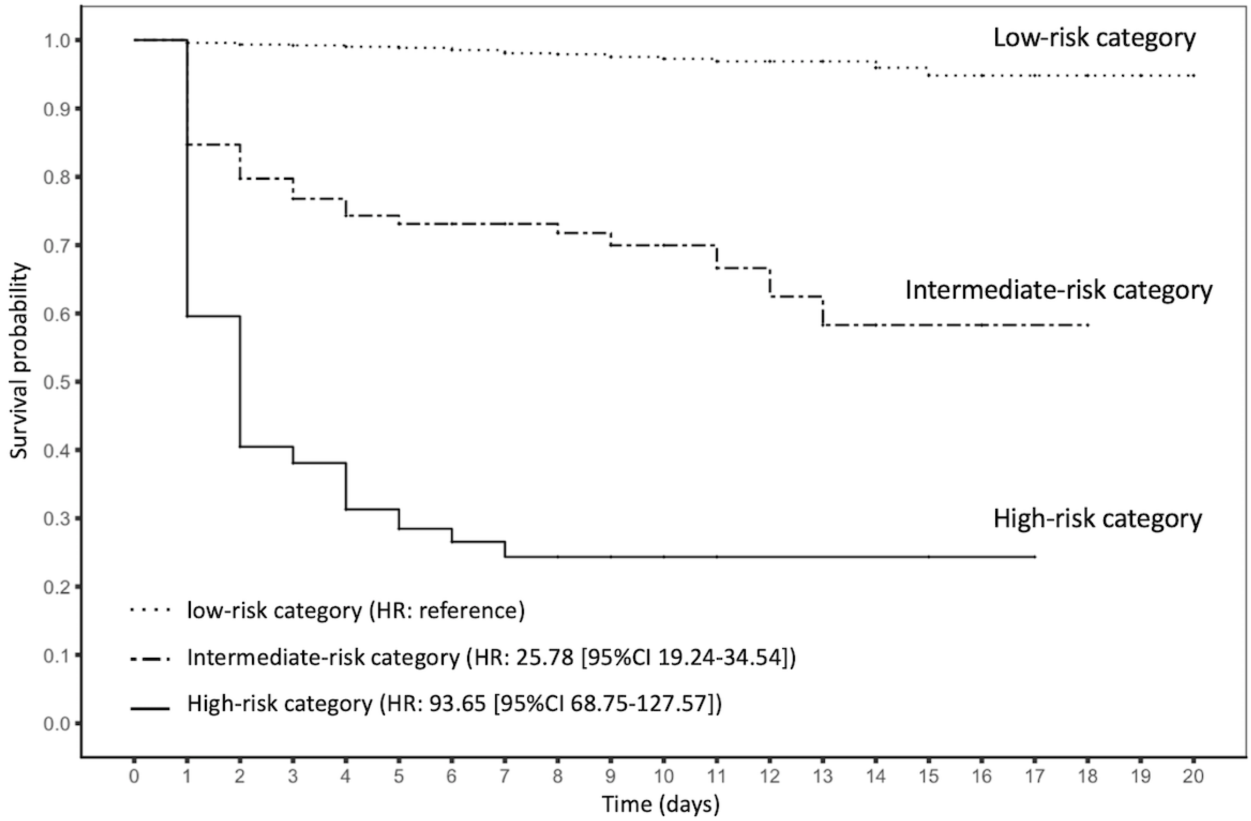
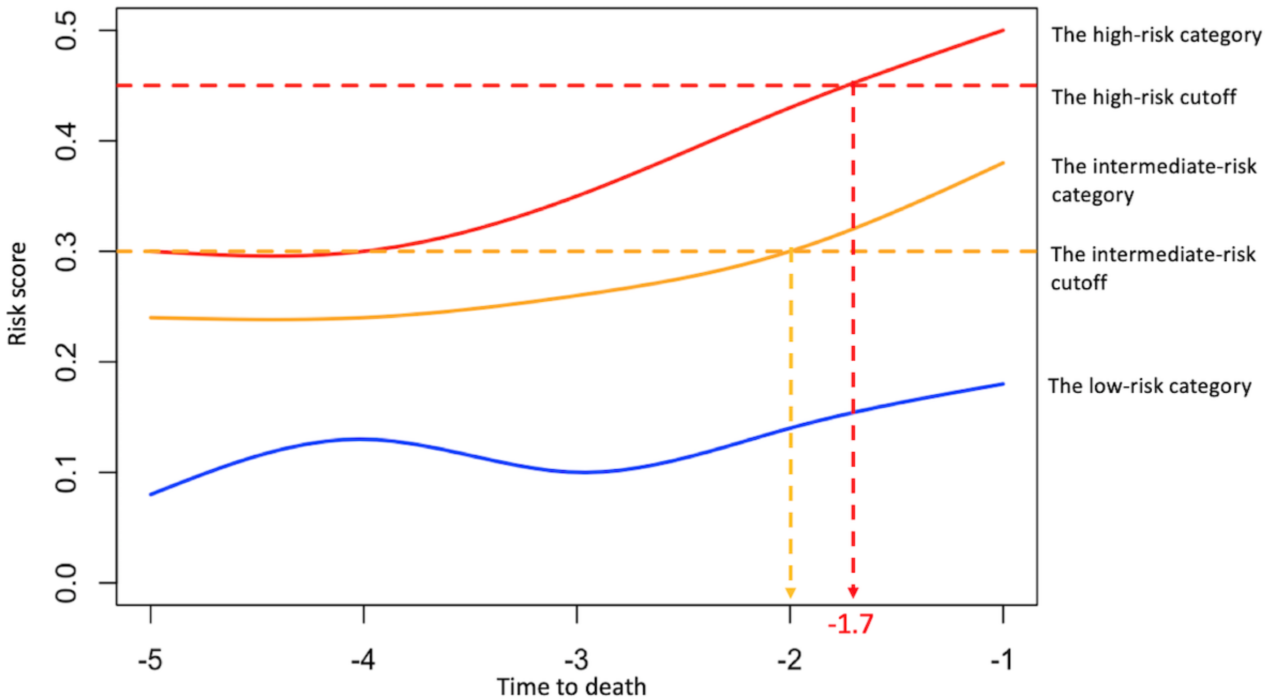



**Figure 4.** The median real-time risk score curves of the patients who passed away stratified by 3 risk categories of the prospective cohort.




### Comparison With Currently Used Methods

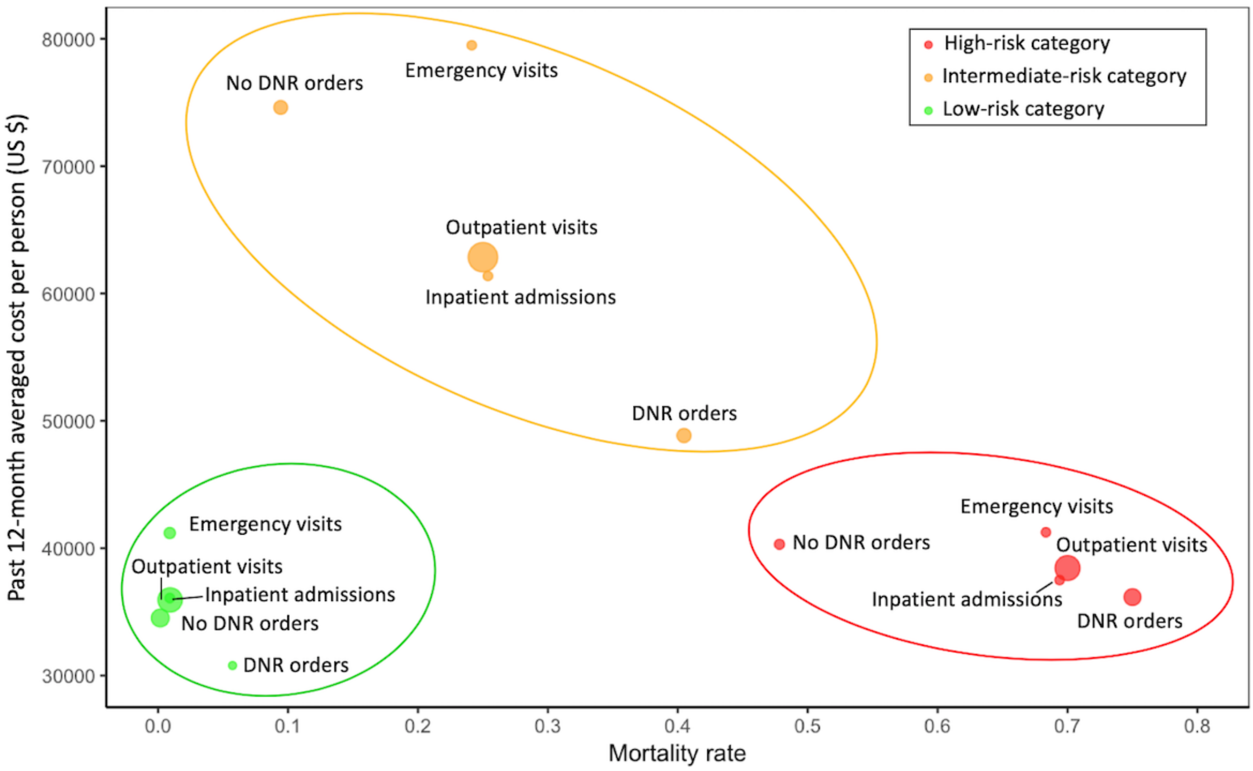
Several EWSs have already been widely used in current hospital care to provide early warnings of clinical deterioration, such as the ViEWS, the National Early Warning Score, and the Modified Early Warning Score [8,9,24-26]. The shared rationale underlying these common EWSs is that a patient's deterioration can be estimated with a numeric score derived from a small number (<10) of core signs of physiological function including, but not limited to, heart rate, breathing rate, body temperature, systolic blood pressure, oxygen saturation, urine output, and level of consciousness. Given that the used parameters have been recognized as vital to life (Multimedia Appendix 5), these EWSs can be readily implemented and are expected to have good predictive ability for life-threatening outcomes. However, by using only vital sign abnormalities, the existing methods attain a high specificity with a low sensitivity [27]. Meanwhile, other innovative EWSs have been implemented with better performance by extracting temporal clinical information from EMRs [13-15,17]. In this study, we hypothesize that integration of the real-time EMR datasets with vital signs, laboratory data, disease diagnosis, and clinical utilization indicators shall lead to an EWS with an improved performance in terms of both sensitivity and specificity. Therefore, we compared the proposed EWS with ViEWS, a well-recognized EWS leader, which was proven to outperform most other systems [9,10]. In this comparison, we applied the abbreviated ViEWS tool on the prospective dataset, which achieved a prospective c-statistic of 0.764, a much lower value than that of the EWS model (c-statistic=0.884; Multimedia Appendix 6). Furthermore, when considering only *high-risk* individuals, the EWS achieved a sensitivity of 26.7% (68/255) and a PPV of 69% (68/99), whereas the ViEWS method attained a much lower sensitivity of 13.7% (35/255) and a PPV of only 35% (35/99). When considering both *high* and *intermediate-risk* patients, the EWS attained a sensitivity of 59.2% (151/255) and a PPV of 35.4% (151/426), which were still much higher than that of ViEWS (a sensitivity of 35.7% (91/255) and a PPV of 21.4% (91/426); Multimedia Appendix 6).

### Impactful Predictors in the Developed Early Warning System

We further adopted the Gini impurity [18] as the indicator of the variable importance, as it usually gives a much faster calculation while providing similar results to the out-of-bag permutation measure. By applying the Gini impurity measurement, we recognized 349 impactful predictors for inpatient mortality from the initial 600 input features. We listed the top 50 most significant features in Multimedia Appendix 7.

Among these features, the proposed EWS recognized several historical clinical utilization features as highly significant predictors of in-hospital deterioration, including ED visits, inpatient admissions, and outpatient visits and clinical costs in the prior 12 months. We grouped patients by the type (ie, emergency, inpatient and outpatient) of their hospital visits and prospectively compared their averaged prior-12-month clinical costs across the 3 determined risk categories, coordinated by their observed inpatient mortality rate (Figure 5). The results showed that these subgroups aggregated naturally into 3 identified risk clusters when plotting by the dimensions of historical clinical costs and observed mortality rates. For patients estimated as *high risk* of inpatient mortality, subgroups of emergency, inpatient, or outpatient encounters, all had higher observed mortality rates but lower clinical costs than those of the *intermediate-risk* patients. On the contrary, *intermediate-risk* patients had dramatic increase of their prior-12-month clinical costs, especially for patients with emergency visits who ended up with a modest rate of inpatient mortality.

**Figure 5.** The averaged prior-12-month clinical costs of distinct clinical utilization subgroups, coordinated by their observed mortality rates. Those subgroups are naturally clustered into 3 mortality risk categories of the prospective cohort. Size of each ball: the median of each group. DNR: do-not-resuscitate.



Furthermore, when focusing on these top impactful chronic-based risk factors, we found that only the diagnoses of cardiovascular diseases, congestive heart failure, or renal diseases were still significantly associated with the mortality outcome, whereas other chronic-based features failed to attain significance in terms of odds ratios (ORs) after applying the propensity score matching analysis in our study (Multimedia Appendix 8). The results of the propensity score matching analysis revealed the insignificant independent effects of some targeted chronic risk factors when matched with other significant risk factors. Therefore, we reason that, in the hospital inpatient mortality setting, instead of being the causality of the mortality outcome, some high-weight chronic-based risk factors could be causally related to the acute setting risk factors or interact with other risk factors (such as demographic characteristics), indirectly and interactively contributing to the prediction of the targeted mortality outcome.

### Patients With and Without a Do-Not-Resuscitate Order

We further investigated the EWS model's discriminative ability in different subgroups of patients with specific diagnoses and conditions (Multimedia Appendix 9), especially patients with and without DNR orders. As confirmed in the validation results, the DNR order patients usually had a much higher inpatient mortality rate than that of the non-DNR order ones (Multimedia Appendix 10). Meanwhile, when looking only at the DNR-order encounters, their mortality rate was still stratified by the 3 distinct risk categories of the EWS; the mortality rate of DNR-order encounters reached its highest value of 75% (57/76) in the *high-risk* category, dropped to 40.5% (68/168) in the *intermediate-risk* category, and plunged to 5.73% (88/1,537) in the *low-risk* category (Multimedia Appendix 10 and Multimedia Appendix 4). This implied that even though some encounters were coded by DNR orders, they still varied significantly in their current in-hospital mortality risk.

## Discussion

### Summary of Principal Findings

In this study, we developed and prospectively validated a real-time EMR-based EWS of inpatient mortality, which predicted encounters' daily and longitudinal probability of inpatient mortality. With a total of 11,762 hospitalized encounters at the prospective validation stage, this model achieved a c-statistic of 0.884, prognosticated *high risk* of death for 99 encounters during their inpatient stay. For these *high-risk* encounters, 40% (40/99) were confirmed to have passed in the subsequent 24 hours, and 69% (68/99) were confirmed to have passed within 7 days after the notification, resulting in their mortality HR as high as 93.65 (95% CI 68.75-127.57) compared with that of the *low-risk* category. Furthermore, the EWS model successfully prognosticated the death of the top 13.3% (34/255) of the dead patients at least 1.7 days before their death.

In this study, we compared the EWS with the well-recognized EWS tool, ViEWS, and demonstrated that the EWS attained a much higher sensitivity and PPV when giving alerts for the *high-risk* patients. Compared with these existing EWSs, the proposed model involved not only traditional predictors of inpatient mortality, such as vital signs and laboratory data [8,9,28,29], but also valuable historical medical features, such as certain disease diagnoses and clinical utilization indicators, which were usually not included in most previous studies [8,9,14]. However, these inpatient setting features, representing patients' baseline differences, can contribute indirectly to patients' distinct hospital mortality rate assessment [30]. In this study, the *intermediate-risk* population in our study, instead of the *high-risk* group, was found to have the highest historical medical costs (Figure 5). This may imply that some of these *high-risk* patients were already coded with DNR orders, directly reducing their clinical costs; others may have deteriorated too rapidly from a healthy status and therefore, never received adequate medical service before death, also resulting in low costs. Therefore, we believe that such historical information in EMR datasets are valuable sources of predictors of inpatient hospital mortality. These risk predictors may interact with other features to facilitate the identification of more true-positive patients, resulting in an improved sensitivity.

### Implications of the Developed Early Warning System

In this study, random forest outperformed other commonly used algorithms on the prospective cohort. As an ensemble tree-based method, random forest has been proven to have high accuracy as it overcomes overfitting by selecting random subsets of features to build smaller trees and is able to handle potential errors caused by unbalanced case-control datasets (in this case, inpatient mortality, where only a relatively small proportion of patients suffered in-hospital death) [31]. In addition, random forest makes no assumptions regarding the predictor features' distributions and correlations and is able to capture features with weak effects as well as their high-level interactions, thus making it suitable for our EMR-based prediction based on multiple correlated covariates [32]. Along with the massively increased data, another well-recognized method, deep learning, is popularly used because of the recent breakthroughs in algorithm development. However, deep learning does not necessarily perform better than linear and nonlinear machine learning methods, as it usually returns a result that is difficult to interpret for domain specialists, and it is more computationally consuming and expensive, especially in the model development stage [33].

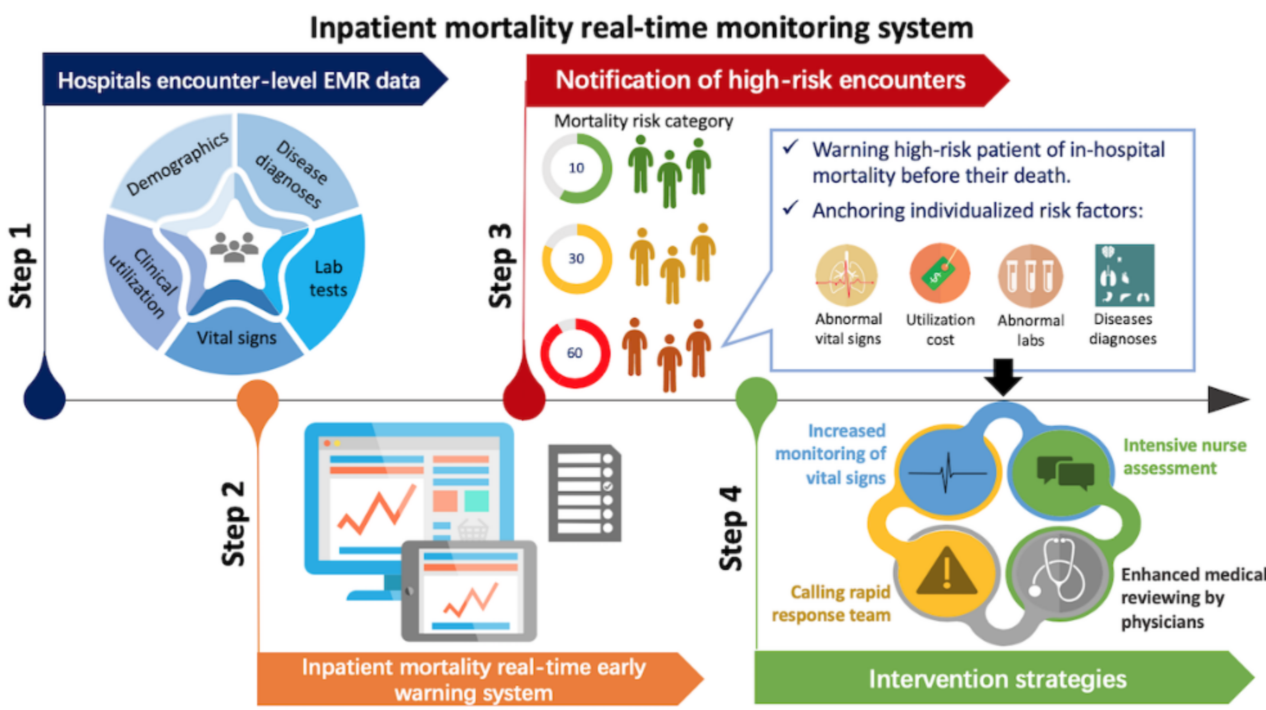
It is worth noting that in the prospective cohort, 31 of the 99 patients who were given alerts for *high risk* of inpatient mortality survived through the entire hospital encounter. After investigating those recovered patients, we found that most of them (25/31) received diagnoses of either cardiovascular diseases, renal disease, cancer, lung disease, or acute cerebrovascular disease, which implied severe acute or chronic disease conditions as well as the requirement of more intense care during their hospital stay. In such an early warning context, it is demonstrated elsewhere that sensitivity and PPV are always considered important indicators; however, these patients who were alerted as high risk yet later recovered may not necessarily be treated as falsely alarmed individuals, as caregivers could always provide clinical intervention or treatment to these high-risk patients during their deterioration process and potentially prevent their death events from occurring [34]. Thus, from the perspective of inpatient mortality reduction and clinical care promotion, it would be valuable to track and summarize those efficient interventions or treatments provided to these high-risk but recovered patients, facilitating evidence-based clinical decision making and individualized care planning for other high-risk patients.

Many of the diseases currently being treated in the wards are major injuries, and these patients could become potential confounders when predicting the hospital inpatient mortality. However, patients with such major injuries are often difficult to define in the EMR system, as there is an issue with the preciseness of their diagnoses when using the International Classification of Diseases (ICD) codes. In this study, taking fracture as an example, we found that when using the standard ICD-10 definitions (Multimedia Appendix 11), the proportion of patients with a fracture diagnosis was relatively small in both the overall cohort and the *high-risk* category (ie, overall: 5.42% [639/11,762]; *high-risk* category: 8.10% [8/99]) and the OR was also not significantly different between cases and controls (OR 1.17; 95% CI 0.67-1.89). To address and verify the impact of major injuries as confounding factors, we hypothesized that young patients are more likely to die of major injuries, whereas the older patients mainly suffered from other severe conditions, and thus, we used age as a summarizing indicator of major injuries in our dataset. After investigating the age-stratified mortality across the identified risk categories (*high*, *intermediate*, and *low*), we revealed that instead of young patients, most true positives in the identified *high* and *intermediate-risk* categories were older than 60 years (Multimedia Appendix 11), who were less likely to die of accidents or major injuries. Therefore, we concluded that in our study cohort, major injuries did not have a significant impact on the inpatient mortality prediction, but we should be careful to consider this confounding effect for applications in the future.

## Utilization and Benefits of the Early Warning System

Previous studies have developed specific in-hospital mortality models suitable for a certain disease or condition, such as acute myocardial infarction [35,36] and congestive heart failure [37]. Compared with these models, the EWS model can be universally applied to all hospitalized patients without restricting them to a certain disease diagnosis. To assist clinical decision making, it can automatically send notifications to physicians and RRT when patients exceed the high-risk threshold, offering a chance at earlier detection of acute events. Furthermore, we can provide clinicians with the real-time risks and specific alerts of the impactful risk factors that the deteriorating patient has and give clinicians suggestions of individualized follow-up health care plans, such as increased monitoring of vital signs, intensive nurse assessments of the patient's condition, and enhanced medical review by physicians [38] (Figure 6).

**Figure 6.** The implementation framework and workflow of the real-time early warning system (EWS), demonstrated in 4 steps: (1) import patient encounters' electronic medical record (EMR) data into the EWS, (2) monitor their inpatient mortality risk scores every 15 min in the user interface after the deployment, (3) use predetermined thresholds to predict the encounters with high risk and intermediate risk of inpatient mortality in a real-time scenario, and (4) highlight or pop up individualized impactful risk factors to help design and implement the subsequent individualized intervention.



Compared with most studies focusing only on patients without DNR orders, the EWS targeted all hospitalized patients regardless of their status of DNR orders. In our study, we found that patients with DNR orders can still be differentiated in terms of their inpatient mortality risk (Multimedia Appendix 10). Previous studies also observed that hospitals' DNR rates could influence the inpatient mortality outcomes in different ways [39]. Therefore, DNR orders do not necessarily indicate the imminent death of the patients in hospital, and the early warning of their death event is important to help the palliative care providers offer supportive services for both patients and their families, such as relieving patients from the symptoms and stress of the illness and letting the family prepare for the deathbed farewell and bereavement. On the other hand, when considering non–DNR-order patients, the identification of their high-risk status could trigger an early warning, activate in-hospital RRTs to a more intensive intervention, and provide a chance to reduce the death or cardiac arrest rate.

In previous studies, limited evidence has been provided to support the conclusion that EWSs have a straightforward effect on the reduction of mortality and cardiac arrests [27,40,41]. With the deployment of the EWS in the BHS hospitals, we will investigate the EWS's long-term benefit on patient health and resource utilization outcomes.

### Limitations

The proposed EWS is built on the EMR data from hospitals located in a relatively small region, and thus the model may not be directly applied to other regions and clinical settings. However, we established the framework and detailed workflow for the construction and validation of the EMR-based inpatient mortality EWS, which can easily be migrated to much broader settings and bigger datasets. In addition, we also consider patient-level social determinants as important and potential data source for in-hospital mortality prediction as most of them are long-term prognosis factors influencing the mortality outcome. Therefore, incorporating such data in the future will make the next-generation EWS model more compelling and robust.

### Conclusions

In this study, by using modern machine learning algorithms, we have developed and prospectively validated an EWS for forecasting inpatient mortality based on patients' EMR data. This EWS prospectively achieved a high predictive accuracy in the validation stage. As a real-time surveillance system that will be integrated into the target medical facilities to assist clinical decision making in the near future, the EWS could trigger an early notification for the patients at high risk of in-hospital mortality, thereby letting clinicians initiate intensive care before the acute event and provide a chance of individualized management to improve the quality of health care.


## Acknowledgments

The authors thank and express their gratitude to the medical practices, physicians, and nurses participating in two acute hospitals within the BHS. They thank the biostatistics colleagues at the Department of Health Research and Policy for critical discussions. They also thank the All Vista Healthcare Center, Ministry of Science and Technology, Taiwan for the support. CY is also supported by the National Natural Science Foundation of China (award #81402762).


## Authors' Contributions

CY, OW, ML, LZ, MX, SH, BJ, HJ, and CZ carried out the initial analysis and interpretation of data and drafted the initial manuscript. FS, KGS, EW, DM, and XL conceptualized and designed the study and critically reviewed and revised the manuscript. GE and DB coordinated and supervised data acquisition and critically reviewed and revised the manuscript. All authors have read and approved this submission for publication. All authors have agreed to be accountable for all aspects of the work.

## Conflicts of Interest

KGS, EW, and XBL are cofounders and equity holders of HBI Solutions, Inc, which is currently developing predictive analytics solutions for health care organizations. The research and research results are not, in any way, associated with Stanford University. There are no patents, further products in development, or marketed products to declare. The remaining authors have no conflicts of interest to declare.

## Multimedia Appendix 1

Summarized demographics and baseline characteristics.

[DOCX File, 17KB - jmir_v21i7e13719_app1.docx ]

## Multimedia Appendix 2

The receiver operating characteristic curves of various algorithms at the prospective validation stage.

[DOCX File, 225KB - jmir_v21i7e13719_app2.docx ]

### Multimedia Appendix 3

The performance of the inpatient mortality early warning system on the prospective cohort, summarized in inpatient-day level positive predictive value, sensitivity, specificity, and relative risk.

[DOCX File, 13KB - jmir_v21i7e13719_app3.docx ]

### Multimedia Appendix 4

The patient distribution in the 3 risk categories identified by the early warning system in the prospective validation cohort.

[DOCX File, 13KB - jmir_v21i7e13719_app4.docx ]

### Multimedia Appendix 5

Summary of variables used in various currently used early warning systems.

[DOCX File, 33KB - jmir_v21i7e13719_app5.docx ]

### Multimedia Appendix 6

The performance comparison between the early warning system model and the well-recognized VitalPAC Early Warning Score method on the prospective dataset.

[DOCX File, 84KB - jmir_v21i7e13719_app6.docx ]

### Multimedia Appendix 7

The top 50 most important features contributed to the inpatient mortality early warning system.

[DOCX File, 21KB - jmir_v21i7e13719_app7.docx ]

### Multimedia Appendix 8

Odds ratios of the impactful chronic-based predictors before and after the propensity score matching analysis.

[DOCX File, 15KB - jmir_v21i7e13719_app8.docx ]

### Multimedia Appendix 9

The observed mortality rates in distinct patient subgroups, stratified by the low-risk (blue), intermediate-risk (yellow), and high-risk (red) categories identified by the early warning system on the prospective cohort.

[DOCX File, 150KB - jmir_v21i7e13719_app9.docx ]

### Multimedia Appendix 10

The inpatient mortality rate of do-not-resuscitate (DNR)-order (orange) and non–DNR-order (blue) populations in the 3 risk categories of the prospective cohort.

[DOCX File, 121KB - jmir_v21i7e13719_app10.docx ]

### Multimedia Appendix 11

Age-stratified mortality across the identified risk categories.

[DOCX File, 117KB - jmir_v21i7e13719_app11.docx ]

### References

1. Chan PS, Khalid A, Longmore LS, Berg RA, Kosiborod M, Spertus JA. Hospital-wide code rates and mortality before and after implementation of a rapid response team. J Am Med Assoc 2008 Dec 3;300(21):2506-2513. [doi: 10.1001/jama.2008.715] [Medline: 19050194]
2. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and early warning systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. Cochrane Database Syst Rev 2007 Jul 18(3):CD005529. [doi: 10.1002/14651858.CD005529.pub2] [Medline: 17636805]

3. Berwick DM, Calkins DR, McCannon CJ, Hackbarth AD. The 100,000 lives campaign: setting a goal and a deadline for improving health care quality. J Am Med Assoc 2006 Jan 18;295(3):324-327. [doi: 10.1001/jama.295.3.324] [Medline: 16418469]
4. Jones DA, DeVita MA, Bellomo R. Rapid-response teams. N Engl J Med 2011 Jul 14;365(2):139-146. [doi: 10.1056/NEJMra0910926] [Medline: 21751906]
5. Forero R, McDonnell G, Gallego B, McCarthy S, Mohsin M, Shanley C, et al. A literature review on care at the end-of-life in the emergency department. Emerg Med Int 2012;2012:486516 [FREE Full text] [doi: 10.1155/2012/486516] [Medline: 22500239]
6. Armitage M, Eddleston J, Stokes T, Guideline Development Group at the NICE. Recognising and responding to acute illness in adults in hospital: summary of NICE guidance. Br Med J 2007 Aug 4;335(7613):258-259 [FREE Full text] [doi: 10.1136/bmj.39272.679688.47] [Medline: 17673769]
7. Opio MO, Nansubuga G, Kellett J. Validation of the VitalPAC™ early warning score (ViEWS) in acutely ill medical patients attending a resource-poor hospital in sub-Saharan Africa. Resuscitation 2013 Jun;84(6):743-746. [doi: 10.1016/j.resuscitation.2013.02.007] [Medline: 23438452]
8. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the national early warning score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation 2013 Apr;84(4):465-470. [doi: 10.1016/j.resuscitation.2012.12.016] [Medline: 23295778]
9. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS--towards a national early warning score for detecting adult inpatient deterioration. Resuscitation 2010 Aug;81(8):932-937. [doi: 10.1016/j.resuscitation.2010.04.014] [Medline: 20637974]
10. Kellett J, Kim A. Validation of an abbreviated Vitalpac™ early warning score (ViEWS) in 75,419 consecutive admissions to a Canadian regional hospital. Resuscitation 2012 Mar;83(3):297-302. [doi: 10.1016/j.resuscitation.2011.08.022] [Medline: 21907689]
11. Sacco Casamassima MG, Salazar JH, Papandria D, Fackler J, Chrouser K, Boss EF, et al. Use of risk stratification indices to predict mortality in critically ill children. Eur J Pediatr 2014 Jan;173(1):1-13. [doi: 10.1007/s00431-013-1987-6] [Medline: 23525543]
12. Moonesinghe SR, Mythen MG, Das P, Rowan KM, Grocott MP. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: qualitative systematic review. Anesthesiology 2013 Oct;119(4):959-981. [doi: 10.1097/ALN.0b013e3182a4e94d] [Medline: 24195875]
13. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*. Crit Care Med 2014 Apr;42(4):841-848 [FREE Full text] [doi: 10.1097/CCM.0000000000000038] [Medline: 24247472]
14. Rothman MJ, Rothman SI, Beals 4th J. Development and validation of a continuous measure of patient condition using the electronic medical record. J Biomed Inform 2013 Oct;46(5):837-848 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.011] [Medline: 23831554]
15. Tabak YP, Sun X, Nunez CM, Johannes RS. Using electronic health record data to develop inpatient mortality predictive model: acute laboratory risk of mortality score (ALaRMS). J Am Med Inform Assoc 2014;21(3):455-463 [FREE Full text] [doi: 10.1136/amiajnl-2013-001790] [Medline: 24097807]
16. Wong J, Taljaard M, Forster AJ, Escobar GJ, van Walraven C. Derivation and validation of a model to predict daily risk of death in hospital. Med Care 2011 Aug;49(8):734-743. [doi: 10.1097/MLR.0b013e318215d266] [Medline: 21478775]
17. Cai X, Perez-Concha O, Coiera E, Martin-Sanchez F, Day R, Roffe D, et al. Real-time prediction of mortality, readmission, and length of stay using electronic health record data. J Am Med Inform Assoc 2016;23(3):553-561. [doi: 10.1093/jamia/ocv110] [Medline: 26374704]
18. Breiman L, Cutler A. The Comprehensive R Archive Network. 2018 Mar 25. Package 'RandomForest' URL: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf [accessed 2019-06-11] [WebCite Cache ID 781QCayMU]
19. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD'16; August 13-17, 2016; San Francisco, California p. 785-794. [doi: 10.1145/2939672.2939785]
20. Ridgeway G. R Project. 2019. gbm: Generalized Boosted Regression Models URL: https://cran.r-project.org/web/packages/gbm/index.html [accessed 2019-06-11] [WebCite Cache ID 781QSUtmm]
21. Karatzoglou A, Smola A, Hornik K. The Comprehensive R Archive Network. 2018. kernlab: Kernel-Based Machine Learning Lab URL: https://cran.r-project.org/web/packages/kernlab/ [accessed 2019-06-11] [WebCite Cache ID 781Pppx4w]
22. Friedman J, Hastie T, Tibshirani R, Simon N, Narasimhan B. The Comprehensive R Archive Network. 2019. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models URL: https://cran.r-project.org/web/packages/glmnet/glmnet.pdf [accessed 2019-06-11] [WebCite Cache ID 781Qfckq7]
23. Ripley BD. The R Language - Seminar for Statistics. 2019. k-Nearest Neighbour Classification URL: https://stat.ethz.ch/R-manual/R-devel/library/class/html/knn.html [accessed 2019-06-11] [WebCite Cache ID 781RvvHOQ]
24. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified early warning score in medical admissions. QJM 2001 Oct;94(10):521-526. [doi: 10.1093/qjmed/94.10.521] [Medline: 11588210]

25. Paterson R, MacLeod DC, Thetford D, Beattie A, Graham C, Lam S, et al. Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit. Clin Med (Lond) 2006;6(3):281-284 [FREE Full text] [doi: 10.7861/clinmedicine.6-3-281] [Medline: 16826863]
26. Wheeler I, Price C, Sitch A, Banda P, Kellett J, Nyirenda M, et al. Early warning scores generated in developed healthcare settings are not sufficient at predicting early mortality in Blantyre, Malawi: a prospective cohort study. PLoS One 2013;8(3):e59830 [FREE Full text] [doi: 10.1371/journal.pone.0059830] [Medline: 23555796]
27. Smith ME, Chiovaro JC, O'Neil M, Kansagara D, Quiñones AR, Freeman M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. Ann Am Thorac Soc 2014 Nov;11(9):1454-1465. [doi: 10.1513/AnnalsATS.201403-102OC] [Medline: 25296111]
28. Asadollahi K, Hastings IM, Gill GV, Beeching NJ. Prediction of hospital mortality from admission laboratory data and patient age: a simple model. Emerg Med Australas 2011 Jun;23(3):354-363. [doi: 10.1111/j.1742-6723.2011.01410.x] [Medline: 21668723]
29. Froom P, Shimoni Z. Prediction of hospital mortality rates by admission laboratory tests. Clin Chem 2006 Feb;52(2):325-328 [FREE Full text] [doi: 10.1373/clinchem.2005.059030] [Medline: 16449218]
30. Hall MJ, Levant S, DeFrances CJ. Trends in inpatient hospital deaths: national hospital discharge survey, 2000-2010. NCHS Data Brief 2013 Mar(118):1-8 [FREE Full text] [Medline: 23742820]
31. Caruana R, Karampatziakis N, Yessenalina A. An Empirical Evaluation of Supervised Learning in High Dimensions. In: Proceedings of the 25th International Conference on Machine learning. 2008 Presented at: ICML'08; July 5-9, 2008; Helsinki, Finland p. 96-103. [doi: 10.1145/1390156.1390169]
32. Liaw A, Wiener M. The R Project for Statistical Computing. 2002. Classification and regression by randomForest URL: https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf
33. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
34. Capan M, Hoover S, Miller KE, Pal C, Glasgow JM, Jackson EV, et al. Data-driven approach to early warning score-based alert management. BMJ Open Qual 2018;7(3):e000088 [FREE Full text] [doi: 10.1136/bmjoq-2017-000088] [Medline: 30167470]
35. McNamara RL, Kennedy KF, Cohen DJ, Diercks DB, Moscucci M, Ramee S, et al. Predicting in-hospital mortality in patients with acute myocardial infarction. J Am Coll Cardiol 2016;68(6):626-635 [FREE Full text] [doi: 10.1016/j.jacc.2016.05.049] [Medline: 27491907]
36. Granger CB, Goldberg RJ, Dabbous O, Pieper KS, Eagle KA, Cannon CP, Global Registry of Acute Coronary Events Investigators. Predictors of hospital mortality in the global registry of acute coronary events. Arch Intern Med 2003 Oct 27;163(19):2345-2353. [doi: 10.1001/archinte.163.19.2345] [Medline: 14581255]
37. Lim E, Cheng Y, Reuschel C, Mbowe O, Ahn HJ, Juarez DT, et al. Risk-adjusted in-hospital mortality models for congestive heart failure and acute myocardial infarction: value of clinical laboratory data and race/ethnicity. Health Serv Res 2015 Aug;50(Suppl 1):1351-1371 [FREE Full text] [doi: 10.1111/1475-6773.12325] [Medline: 26073945]
38. Mathukia C, Fan W, Vadyak K, Biege C, Krishnamurthy M. Modified early warning system improves patient safety and clinical outcomes in an academic community hospital. J Community Hosp Intern Med Perspect 2015;5(2):26716 [FREE Full text] [doi: 10.3402/jchimp.v5.26716] [Medline: 25846353]
39. Walkey AJ, Weinberg J, Wiener RS, Cooke CR, Lindenauer PK. Association of do-not-resuscitate orders and hospital mortality rate among patients with pneumonia. JAMA Intern Med 2016 Jan;176(1):97-104. [doi: 10.1001/jamainternmed.2015.6324] [Medline: 26658673]
40. McNeill G, Bryden D. Do either early warning systems or emergency response teams improve hospital patient survival? A systematic review. Resuscitation 2013 Dec;84(12):1652-1667. [doi: 10.1016/j.resuscitation.2013.08.006] [Medline: 23962485]
41. Alam N, Hobbelink EL, van Tienhoven AJ, van de Ven PM, Jansma EP, Nanayakkara PW. The impact of the use of the early warning score (EWS) on patient outcomes: a systematic review. Resuscitation 2014 May;85(5):587-594. [doi: 10.1016/j.resuscitation.2014.01.013] [Medline: 24467882]


## Abbreviations

**BHS:** Berkshire Health System
**DNR:** do-not-resuscitate
**ED:** emergency department
**EMR:** electronic medical record
**EWS:** early warning system
**HR:** hazard ratio
**ICD:** International Classification of Diseases
**OR:** odds ratio
**PPV:** positive predictive value

**ROC:** receiver operating characteristic
**RRT:** rapid response team
**ViEWS:** VitalPAC Early Warning Score

*Edited by G Eysenbach; submitted 15.02.19; peer-reviewed by G Tognola, Y Chu, GE Iyawa; comments to author 20.03.19; revised version received 08.05.19; accepted 25.05.19; published 05.07.19.*

*Please cite as:*
*Ye C, Wang O, Liu M, Zheng L, Xia M, Hao S, Jin B, Jin H, Zhu C, Huang CJ, Gao P, Ellrodt G, Brennan D, Stearns F, Sylvester KG, Widen E, McElhinney DB, Ling X*
*A Real-Time Early Warning System for Monitoring Inpatient Mortality Risk: Prospective Study Using Electronic Medical Record Data*
*J Med Internet Res 2019;21(7):e13719*
*URL: https://www.jmir.org/2019/7/e13719/*
*doi:10.2196/13719*
*PMID:31278734*

Original Paper

# Implementation of a Digitally Enabled Care Pathway (Part 2): Qualitative Analysis of Experiences of Health Care Professionals

Alistair Connell[1,2], MBBS; Georgia Black[3], PhD; Hugh Montgomery[1], MD; Peter Martin[3], PhD; Claire Nightingale[3,4], PhD; Dominic King[2], PhD; Alan Karthikesalingam[2], PhD; Cían Hughes[2], MSc; Trevor Back[2], PhD; Kareem Ayoub[2], MA; Mustafa Suleyman[2]; Gareth Jones[5], PhD; Jennifer Cross[5], PhD; Sarah Stanley[5], MSc; Mary Emerson[5], BSc; Charles Merrick[5]; Geraint Rees[6], PhD; Christopher Laing[5], MD; Rosalind Raine[3], PhD

[1]Centre for Human Health and Performance, University College London, London, United Kingdom
[2]DeepMind Health, London, United Kingdom
[3]Department of Applied Health Research, University College London, London, United Kingdom
[4]Population Health Research Institute, St. George's, University of London, London, United Kingdom
[5]Royal Free London NHS Foundation Trust, London, United Kingdom
[6]Faculty of Life Sciences, University College London, London, United Kingdom

**Corresponding Author:**
Rosalind Raine, PhD
Department of Applied Health Research
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 (0) 20 7679 1713
Email: r.raine@ucl.ac.uk

## Abstract

**Background:** One reason for the introduction of digital technologies into health care has been to try to improve safety and patient outcomes by providing real-time access to patient data and enhancing communication among health care professionals. However, the adoption of such technologies into clinical pathways has been less examined, and the impacts on users and the broader health system are poorly understood. We sought to address this by studying the impacts of introducing a digitally enabled care pathway for patients with acute kidney injury (AKI) at a tertiary referral hospital in the United Kingdom. A dedicated clinical response team—comprising existing nephrology and patient-at-risk and resuscitation teams—received AKI alerts in real time via Streams, a mobile app. Here, we present a qualitative evaluation of the experiences of users and other health care professionals whose work was affected by the implementation of the care pathway.

**Objective:** The aim of this study was to qualitatively evaluate the impact of mobile results viewing and automated alerting as part of a digitally enabled care pathway on the working practices of users and their interprofessional relationships.

**Methods:** A total of 19 semistructured interviews were conducted with members of the AKI response team and clinicians with whom they interacted across the hospital. Interviews were analyzed using inductive and deductive thematic analysis.

**Results:** The digitally enabled care pathway improved access to patient information and expedited early specialist care. Opportunities were identified for more constructive planning of end-of-life care due to the earlier detection and alerting of deterioration. However, the shift toward early detection also highlighted resource constraints and some clinical uncertainty about the value of intervening at this stage. The real-time availability of information altered communication flows within and between clinical teams and across professional groups.

**Conclusions:** Digital technologies allow early detection of adverse events and of patients at risk of deterioration, with the potential to improve outcomes. They may also increase the efficiency of health care professionals' working practices. However, when planning and implementing digital information innovations in health care, the following factors should also be considered: the provision of clinical training to effectively manage early detection, resources to cope with additional workload, support to manage perceived information overload, and the optimization of algorithms to minimize unnecessary alerts.

***(J Med Internet Res 2019;21(7):e13143)*** doi:10.2196/13143

**KEYWORDS**
nephrology; acute kidney injury

# Introduction

## Background

In many health systems, the ageing demographic of hospital patients is accompanied by worsening health and a greater need for diverse investigations and treatments. As a result, care pathways are increasingly complex and ever more reliant on access to relevant data and on communication between individuals and multidisciplinary teams [1]. In this context, although adverse events—such as acute kidney injury (AKI), cardiac arrest, or clinical decline which necessitates high-dependency care—are commonly a consequence of the natural history of the underlying disease, they may also occur because of delays in treatment pathways [2]. Evidence suggests that patient outcomes improve where such clinical decline is detected and acted upon early, and substantial effort has been made worldwide in this regard, for instance, through the use of *track and trigger* scoring systems to detect decline [3] or the provision of emergency response teams [4,5]. However, such changes have not been matched by other key components of care delivery: the manner in which health care teams communicate and the manner in which data are accessed and presented. Globally, the most widely used hospital communication system continues to be the pager [6], whereas data are accessed from paper records and a range of disparate and disconnected electronic data repositories. Care might be improved if interpersonal communication were enhanced and if it were possible to readily access data in a form that allowed rapid assessment of the patient's status.

The introduction of digital technologies offers one potential solution, allowing ready detection of the deteriorating patient, communication among health care workers, and immediate access to patient data in a user-friendly and appropriate format, thus improving outcomes. The embedding of digital technologies into health care is now a priority in the United Kingdom [7] and internationally [8]. However, the adoption of such technologies into health care has been little studied, and the impacts on users and the broader health system are poorly understood. We sought to address these issues by studying the impacts of introducing a digitally enabled care pathway for patients with AKI.

AKI—a sudden reduction in kidney function diagnosed by changes in serum creatinine [9]—is common, appears across multiple care pathways, and is associated with significantly increased mortality, morbidity, and cost of health care [10-15]. However, substantial deficits exist in all key processes of AKI care including early recognition and therapy, appropriate escalation to specialist or critical care services, and follow-up [16]. In an effort to expedite and standardize diagnosis, the National Health Service (NHS) mandated the use of a new diagnostic algorithm in all English hospitals in 2014 [17] and provided guidance as to how the algorithm could be implemented [18]. However, simple alerting to the presence of AKI does not seem to improve outcome [19]. We therefore designed a new care pathway that encompassed AKI detection, mobile alerting of a dedicated response team comprising multiple specialists, and the provision of protocolized care [20]. Evaluation of the implementation of the digitally enabled care pathway with regard to impacts on processes of care, clinical outcome, and the cost of care delivery are described elsewhere [21,22].

## Objectives

We present a qualitative evaluation of the experiences of users and other health care professionals whose work was affected by the implementation of the new care pathway. We sought to characterize the impacts on staff of such automated alerting, mobile results viewing, and ready communication, with particular focus on their working practices and interprofessional relationships.

# Methods

## Setting

The digitally enabled pathway was designed and implemented at the Royal Free Hospital (RFH)—a large, acute, tertiary referral hospital providing a range of acute services (including a 34-bed intensive treatment unit and an inpatient nephrology service) in central London, United Kingdom. The care pathway has been described in detail elsewhere [20] and is summarized in brief below.

### *The Preimplementation Care Pathway*

Before making the changes to the care pathway described here, pathology results were viewed by ordering clinicians in batch at the end of the working day using desktop computers. A message linking to local Web-based clinical guidelines was appended to any creatinine result suggestive of AKI in the electronic patient record (EPR), and such results were also communicated to the clinical area by biochemistry staff by telephone. In its early stages, AKI was typically managed independently by general acute care and various specialty teams; specialist input was requested through hospital pagers and telephone communication at the discretion of referring clinical teams.

### *The Digitally Enabled Care Pathway*

Streams (DeepMind Technologies Ltd) is a mobile app deployed on iPhone operating system–enabled smartphones. It processes relevant routinely collected clinical and demographic data through secure integration with hospitals' existing information systems; owing to the need for real-time event-driven data, Health Level Seven (version 2, Health Level Seven International) feeds were used for integration with the laboratory information management system and electronic medical record. It was first registered with the Medicines and Healthcare Products Regulatory Agency as a Class I, nonmeasuring, nonsterile medical device under the European Union Medical Device Directive (1993) on August 30, 2016. Future revisions of the device may be classified at a higher level under the new medical device regulation.

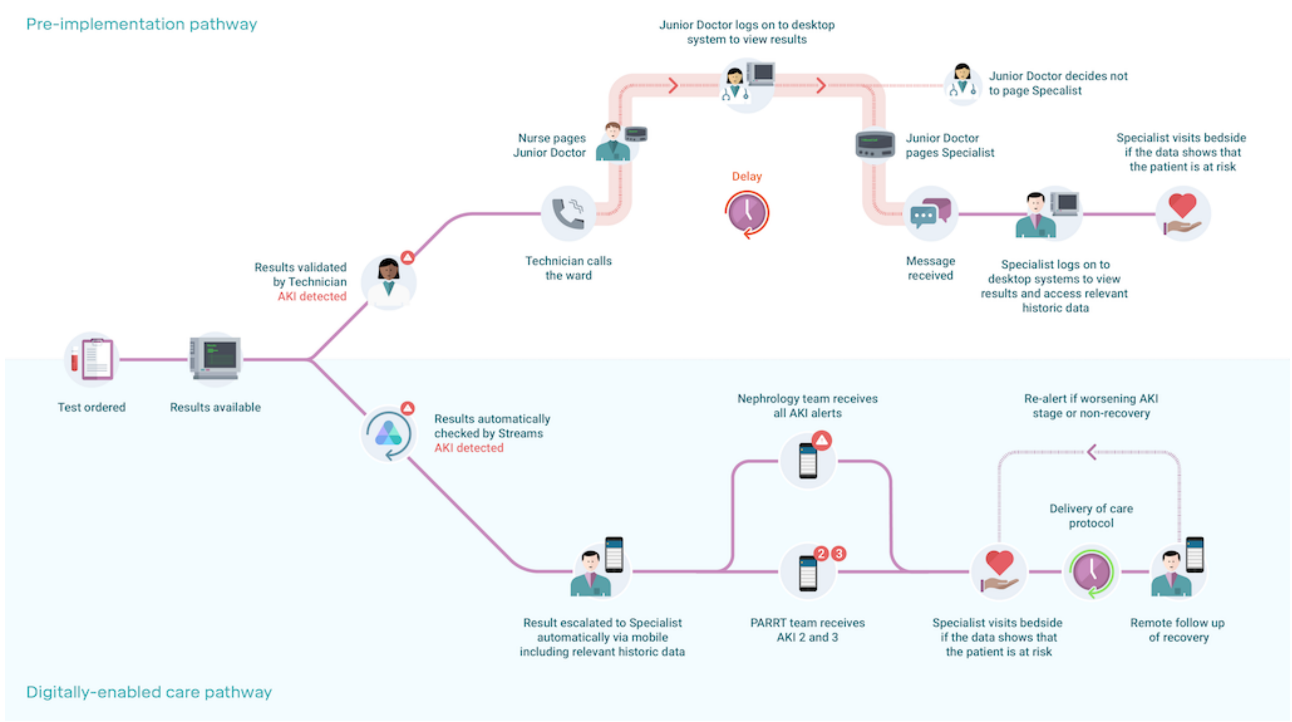
Streams analyzes serum creatinine results immediately and continuously, alerting clinicians in real time to all potential AKI cases as defined by the NHS England AKI algorithm. The app also provides clinicians with data relevant to AKI management, including a graphical trend view of serum creatinine, specific flags for the presence of life-threatening AKI complications (such as hyperkalemia), details of any previous AKI episodes, demographic information, and past medical history from coded Hospital Episode Statistics data. Videos demonstrating Streams functionality can be found on the DeepMind Health support website [23].

AKI alerts are sent in real time to a specialist clinical response team (henceforth, the *AKI response team*), comprising the RFH's existing *patient-at-risk and resuscitation team* (PARRT) and nephrology team. The PARRT (Clinical Nurse Specialists who review at-risk or deteriorating inpatients) receive alerts on all patients with AKI stages 2 and 3 and are on site 24 hours a day. The nephrology team comprises a renal consultant and specialty registrar, both of whom receive all AKI notifications. The registrar is on site 24 hours a day and is typically the first responder. The consultant can triage alerts through secure remote access if off site, providing clinical supervision and subsequent patient review where needed. Through Streams, the AKI response team triages alerts, communicates with other team members, and documents the outcome of clinical reviews. Relevant contacts and clinical guidance are available on Streams phones.

The AKI response team prioritizes patients for review according to the information available in Streams. Patients with life-threatening complications or deemed at high risk are reviewed immediately, whereas a case review within 2 hours is suggested for all other alerts. Upon review, a care protocol (based on existing best practice guidelines [24,25]) is annotated and entered into the patient's paper notes alongside an advisory sticker for key nursing actions (Multimedia Appendix 1).

Although the nephrology team occasionally takes over patient care, overall responsibility for care rests with patients' primary teams. AKI recovery is monitored remotely in-app. Realerting for AKI that has not recovered is enabled 48 hours after the first alert. However, worsening of AKI stage or the development of a new complication (eg, hyperkalemia) at any time results in a further notification. Follow-up reviews are undertaken by the AKI response team according to clinical judgement. A diagram outlining the pre- and postintervention care pathways is provided in Figure 1. Although clinical guidelines and specialist response teams existed before the new care pathway, the goals of implementation of the care pathway were to improve the reliability and speed at which AKI recognition and appropriate specialist review occurred.

**Figure 1.** Pre- and postintervention care pathways. AKI: acute kidney injury.




## Implementation

Before deployment, Streams users attended training events and accessed a video users' guide to both the Streams app and the clinical pathway. Feedback from the AKI response team was gathered during a 16-week pathway optimization period (January-May 2017), during which key adjustments to the Streams app and associated clinical pathway were made. The optimized care pathway has been deployed continuously at RFH since May 2017.

## Data Collection

We undertook an exploratory case study approach to understand the implementation of Streams within a single site. Data collected included in-depth semistructured interviews and brief nonparticipant observations. The interview guide (included in the Multimedia Appendix 1) was developed by the research team, which included physicians with extensive experience in clinical nephrology and intensive care medicine and experts in health services research. Interviews explored the impacts of the care pathway on staff members and the care delivered to patients, with a particular focus on working practices and interprofessional relationships. Face-to-face interviews began 1 month after the start of the pathway optimization period and were spaced throughout a 16-month period of implementation and evaluation (February 2017 to May 2018). Purposive sampling was employed following a key informant strategy [26] that identified individuals with important roles in the environment under study who had expert knowledge to share impartially. A sample size of 20 is typical for a case study such as this, in line with both international consensus guidance and common practice in qualitative research [27-29]. Furthermore, the total number of users involved in providing the care pathway was small, which necessarily restricted the number of interviews that could be conducted. A list of potential respondents was drawn up to ensure representation from both groups in the AKI response team (ie, PARRT and nephrology teams) and clinicians from the wider hospital community affected by the care pathway, as well as a diverse range of clinical experience and level of comfort with mobile technologies. A total of 20 respondents were approached and 19 consented. In total, 8 PARRT nurses (5 *band 7* and 3 *band 8* or above) and 8 nephrologists (4 registrars and 4 consultants) were interviewed from the AKI response team. Of the respondents, 3 (2 consultant physicians and 1 medical registrar) from the wider hospital community were selected as a result of their frequent interactions with the AKI response team. Interviews were conducted by the lead researcher (AC). Each respondent was interviewed once. Respondents were informed that the interviewer was from a university and that the research was independent of the RFH. Field observations were undertaken during the first 4 weeks of the pathway optimization period. For these, the lead researcher observed user behavior in the emergency department and in inpatient wards during day and evening shifts using extensive note-taking to document users' interaction with the Streams app and impacts on working practices and interprofessional relationships.

### Data Analysis

Data were analyzed using a combination of inductive and deductive thematic analysis techniques [30]. First, quotes from each interview were arranged into a matrix (Multimedia Appendix 1) in which rows represented individual respondents and columns represented categories that aligned to the basic principles of the intervention pathway (for example, the triage of AKI alerts). The matrix was populated by 2 researchers (AC and GB) who independently analyzed the entire dataset. Researchers met regularly to critique and challenge each other's allocations; these were in turn reviewed with the lead researcher (RR), a process that enabled us to compare different professional groups' perspectives and identify discordant views. The group then synthesized new descriptive codes based on emergent themes in the matrix (for example, the impact of real-time information availability) and assigned the quotes to these. Discordant quotes that challenged these themes were routinely sought and discussed. Additional independent oversight was provided by the lead researcher who identified additional quotes of relevance and refined the final themes. We employed the principle of *keyness* in our analysis [31], drawing out novel issues that might be generalizable and relevant to the adoption of digital health products in clinical practice (for example, how mobile working tools impact established clinical workflows). Our results present representative quotes for each theme, the titles of which were iterated through the writing process. We therefore took both a descriptive and an interpretive approach to analysis, first understanding how the intervention was used and how it affected clinical practice, then considering the intervention in terms of cultural practices and overarching *meta-themes* about mobile app use.

### Ethical Approval

The digitally enabled care pathway constituted a new standard service at RFH. Plans for the evaluation of the digitally enabled care pathway were independently reviewed by the University College London Joint Research Office. They directed that this study fell under the remit of service evaluation (rather than research), as per guidance from the NHS Health Research Authority [32]. As such, the service evaluation was registered locally with the RFH audit lead and medical director, and no participant consent was required. An independent data monitoring committee (which included a patient member) reviewed all analyses before preparation for publication. A full list of committee members is provided in the Multimedia Appendix 1.

## Results

Although interviews focused on the deployment of a pathway for patients with AKI, they also sought to more broadly characterize the impacts of automated alerting, mobile results viewing, and ready communication on staff, with a particular focus on their working practices and interprofessional relationships. In this respect, 3 central themes emerged.

### Theme 1: The Impact of Real-Time Information Availability

This theme relates to respondents' experiences of the benefits and drawbacks of real-time mobile access to patient data. The app provided automated alerts about patients' kidney function and gave staff access to current and historical information, such as previous pathology and coded diagnoses. These functions were reported to save valuable time among participants from both teams:

> *Being able to look up the blood results for anyone in the hospital wherever you are is unparalleled. [...] it feels almost archaic these days, to go and see a patient and then go and sit down in front of a computer 15 minutes later. As a doctor, you have to integrate what you know about them at the time of seeing them. So if you could literally have this phone, look at the results, go and see them... Or even look at it while you are seeing them. [...] It must save at*

*least - I don't know if you could analyze it - but it must save at least a couple of hours in a day.* [Respondent 3: Nephrology team]

This in turn expedited rapid intervention for deteriorating patients, wherever they were in the hospital:

> *The speed at which it happened was impressive. [...] I happened to be in A&E [Emergency Department] and got the alert of someone with severe kidney injury. [...] The patient was admitted to [...] a specialist renal ward [...] within 2 or 3 hours, which I don't think would have happened without the app. [...] I think it streamlines care and speeds up the time in which they get a specialist renal review.* [Respondent 9: Nephrology team]

> *I personally have noticed [...] patients who have flagged up on the app that the clinical management has been poor up to that point. When we get involved, or the renal team get involved, that management changes [...] It has definitely saved people's lives.* [Respondent 14: PARRT]

> *Being able to access all the bloods for the patients in the hospital and to be able to be alerted to the sick ones and already know about them before we usually do... Sometimes you know about them before the crash bleep comes through. You turn up and you think, "That was actually the alert I was coming to see."* [Respondent 10: PARRT]

Participants in both teams found alerts to be particularly valuable for patients whose lead consultant was not a physician:

> *The most value came from patients under [...] surgical patients, for whom the list of priorities for their clinicians are very different from what [physicians] look for when they are looking after a patient. For those [patients], getting a rapid alert about deranged renal function is very valuable.* [Respondent 6: Nephrology team]

The provision of results and real-time clinical alerts and team communication via mobile phones introduced workload for clinicians in a new modality. Overall, experienced clinicians were able to integrate AKI alerts into their existing duties, discriminating between high- and low-priority cases and using this information to adjust their current priorities:

> *I would intermittently [...] check it, like I would [...] check emails, [...] check it every hour or so, something like that. And within 5 minutes or so I could easily flick through the alerts and [...] identify which ones I needed to see. [...] I felt it was very easy to use, I think some people when they were trying to use it would try and respond immediately to every alert. I wouldn't personally, I didn't think that was the best way to do it. Intermittently checking it throughout the day, I managed to keep on top of things.* [Respondent 9: Nephrology team]

However, some more junior clinicians in both teams felt that the pathway created additional workload and suggested that clinical review might not be deliverable to all patients by the AKI response team as configured currently:

> *It does increase our work. Some days [...] we can have eight or nine referrals. But there is obviously a huge issue about workload for many people. But if we need to increase the size of our team because of this then that's a good thing. And also it highlights [...] the acuity of our patients in our hospital. These patients are not straightforward.* [Respondent 19: PARRT]

Others pointed out that the added burden of the app was related to the volume of *false positive* alerts produced by the mandated NHS AKI algorithm:

> *...if the noise of the system could be reduced it would be a lot better. If [we were] able to get rid of all the nonsense alerts, that would be fantastic.* [Respondent 1: Nephrology team]

Some respondents from the nephrology team pointed out that although important information was now more readily available, this created additional anxiety because it was not clear who was primarily responsible for delivering timely care. Once again, the need to expand the responsive workforce was suggested as a way to mitigate the workload-associated stress:

> *...as Renal [Registrars], [...] you are always now, in the back of your head, thinking "I've got this other job to do." And I think it does create... not anxiety that keeps you up at night... But it's another anxiety when you already have enough anxiety! So I think even if it was available in the hands of more people, or we were a bit clearer that during times of people being unwell who are your own patients, you shouldn't prioritise Streams people because they are under another team, then that's fine. That's one way of dealing with it.* [Respondent 3: Nephrology team]

In summary, although the digitally enabled care pathway was widely valued for increasing efficient access to patient information, thus facilitating better care, some respondents reported increased workload and anxiety.

### Theme 2: The Implications of Early Detection

The digitally enabled care pathway was designed to expedite identification of AKI so as to avoid deterioration and improve outcomes, meaning that staff who normally respond to critically unwell patients would now also attend patients at lower risk. There was a divergence of opinion among members of the AKI response team about the rationale for this approach. Respondents in both teams pointed out that the AKI algorithm identified deteriorating patients at an earlier stage than was possible through other means:

> *It's a good thing from the point of view that I know there are patients that are potentially sick out there. [...] You could have an AKI and look relatively well initially. But [...] nobody would have known about those patients.* [Respondent 8: PARRT]

Others emphasized the benefits of early recognition for patient health and in terms of reducing the complexity of required interventions:

> *I think it does a good job. We pick up patients that would maybe sit for another day or so before we pick them up. It's certainly beneficial. It is more work, but we might be saving ourselves work in a couple of days, we might have to do more stuff to catch up.* [Respondent 15: PARRT]

However, some respondents in the nephrology team did not see the point of being alerted to low-risk patients:

> *...you get a lot of AKI stage 1s. They build up. Looking through those and dismissing them each time is time consuming. The AKI stage 2 and 3 [alerts] are more helpful for me to look at, so I tend to just look at those and dismiss the stage 1s.* [Respondent 18: Nephrology team]

Some respondents in this team also pointed out that early identification could not necessarily be aligned to early intervention because of a lack of knowledge with respect to appropriate management:

> *The patients you definitely need to see are the patients that have acute renal failure with a creatinine of 300 or 400 [µmol/L] that and it's going up - patients you'd normally want to see [...] The other patients [...] that have a rising creatinine, but the creatinine is not very high - it doesn't mean that they don't need to be seen necessarily. We are not trained as doctors to look after those sorts of patients.* [Respondent 1: Nephrology team]

Thus, the shift toward earlier detection highlights the need to consider the resources required to manage both early and late disease and the training needed to enable clinicians to effectively intervene at an earlier stage.

## Theme 3: Behavioral Effects of the Care Pathway

The final theme demonstrates how real-time data provision affected relationships between users of the digital intervention and the broader health system with which the users interacted, together with how these changes impacted upon beliefs, behaviors, and care delivery. These are described in 3 subcategories below.

First, the digital care pathway affected behaviors *within* clinical teams in a number of ways. An immediate benefit was the use of mobile phones for team communication, experienced by members of both teams:

> *I've found the [mobile phone] really useful because I've been able to message my team when I'm out seeing a patient, rather than finding a phone and to bleep them with and waiting for them to answer.* [Respondent 5: PARRT]

However, a disadvantage was also identified: junior members of the nephrology team do not currently use the Streams app. Unequal access to patient data occasionally reversed the usual direction of communication of information from junior doctor to consultant, which some consultants suggested impeded learning opportunities:

> *I think it's important for [Junior Doctors] to understand what the decisions about their patients are. They have to be across the data. And that's why I prefer getting information from them [...] We were in the position where I was telling the Juniors what the blood results were. It makes me uncomfortable and it makes them uncomfortable.* [Respondent 6: Nephrology team]

In addition, the visualization of each other's triage decisions within the app (a feature specifically requested by users) revealed the hitherto unrecognized variations that exist in professional judgements. Respondents in both teams raised the fact that knowledge of others' decisions sometimes confused rather than clarified the clinical decision-making processes of colleagues:

> *I was quite surprised about how other people triaged initially. I felt we'd be much more similar in our thinking, because when we talk about other things we do think similarly about other stuff. [...] I felt like - probably naively - that everyone would do what I did. And they didn't at all.* [Respondent 11: PARRT]

Second, the digital pathway had an impact on relationships *between* clinical teams. Several members of the nephrology team were uncomfortable about providing a clinical opinion when not solicited by the clinicians primarily responsible for that patient's care:

> *So you might [...] call the team and say "we suggest you give some fluid," but I don't think it's ethical to prescribe it yourself. After all, they might say, "Listen, he has heart failure." So you can't intrude.* [Respondent 2: Nephrology team]

However, there was an indication among several members of the PARRT team that this concern might be limited to communication between different specialty doctors:

> *I think the doctors have found it more difficult because in medicine there is this real model of, "[...] I don't see this patient unless I'm asked to see them". There's this formality. [...] Nurses don't think like that, people are used to us showing up. So it's been easier for us to think of every patient that we see as our patient, our problem, our sick patients. I think that's been easier for this team to absorb and deal with.* [Respondent 11: PARRT]

This is relevant because the type of information available also changed the professional group with whom this team directly liaised:

> *[I will] always [speak to] the nursing staff, because you are on their ward. It's only polite and also you will generally be recommending frequencies of obs; they need to know what's going on. And someone from the medical team. I would say the app is making me speak to more senior doctors more. [...] I'd be more likely to seek out a consultant and say "by the*

> *way, this person has alerted" and show them the app.* [Respondent 5: PARRT]

Third, the care pathway had an impact on the relationship between clinicians and their patients. In particular, several PARRT team members described that alerts identified patients at an earlier stage of AKI than was the case through established clinical pathways (eg, monitoring of vital signs). This may have led to an unexpected and evolving role for some members of the team. Several respondents described how the care pathway enabled them to help patients make informed decisions surrounding end-of-life care. For example:

> *Why do we have to talk about end of life just as I'm about to die? [...] We could plan. Every single person we've been referred today has a terminal disease. [...] Trying to move the decision making back, in a more timely way. [...] We are getting an alert before they have even triggered [via vital signs], so we can probably have a sensible conversation with a patient with capacity.* [Respondent 4: PARRT]

## Discussion

### Principal Findings

Qualitative and quantitative results from our mixed-methods service evaluation suggest that the digitally enabled care pathway has positive impacts on patient care [21,22]. Here, we demonstrated the ability to intervene in the treatment of deteriorating patients more quickly and the opportunity for earlier, more constructive end-of-life planning. The ability to integrate mobile results viewing into existing clinical workflows also appears to increase efficiency through the immediate access to specific and contextual information. However, this comes at a price, particularly for some junior staff, in terms of anxiety associated with increasing numbers of *priority* patients and information overload, in part exacerbated by false positive alerts. We also highlight the hitherto unrecognized need to ensure that relevant clinical specialties include training in prevention if we are to optimize the value of digital innovations that promote early intervention. These findings suggest that the true cost-effectiveness of such innovations cannot be assessed until the balance between early intervention leading to better outcomes and increased workload is ascertained. Maximization of utility also requires finding the most appropriate balance between sensitivity and specificity for clinical alerts. This can be difficult to achieve: although it is recognized that the NHS England AKI algorithm produces false positives, some argue that this is a necessary trade-off for enhanced sensitivity [33].

### Strengths and Limitations

This evaluation has a number of strengths. First, it benefits from the diversity of our respondent sample, presenting multiple perspectives on the intervention based on cultural differences between different professionals and teams. Second, the robust analysis uncovered issues that are likely to be generalizable to the implementation of other digital technologies in health care. Finally, our analysis team includes researchers from medical, public health, and psychology backgrounds, creating trustworthiness by encouraging debate and multidisciplinary interpretation.

Our evaluation has a number of limitations. First, we initiated interviews during the *pathway optimization* period to enable us to gather insights that would improve the pathway itself. However, interviews at this early stage may have been more prone to include negative feedback before users were used to the changes described. The period in which interviews were conducted did not allow us to assess whether perceptions of the care pathway changed over time. Second, interviews were conducted in a single clinical setting. Although our methods allowed us to identify the *active ingredients* of a digital intervention in an acute setting with communication systems familiar to health care teams worldwide, the magnitude of the efficiency benefits reported may vary according to the digital maturity of the health care environment studied.

### Comparison With Prior Work

Few studies of AKI alerting systems have been previously described. Alavijeh et al [34] used a questionnaire to assess satisfaction with a new automated AKI warning system among physicians working in primary care. The questionnaire used was limited to questions about respondents' knowledge of the existence of the alert system, perceptions of its utility, and impact on practice, making comparisons with our findings difficult. However, in common with our findings, many respondents found the alert system *valuable*. Kanagasundam et al [35] examined the effects of an interruptive AKI clinical decision support system (embedded within an EPR system) through a series of semistructured interviews with stakeholders. Themes revealed were similar to those encountered in the generic clinical decision support literature, namely, alert fatigue and user dissatisfaction with mandatory interactions. Although respondents in Kanagasundam's evaluation believed that the alerts led to earlier patient assessment, some clinicians found them to be an *insult to their knowledge*. A major reason for dismissing alerts was users' need to review a comprehensive dataset (including historical creatinine) at the point of alert. We overcame this limitation by including a curated summary of relevant clinical data in-app at the point of alert. Kanagasundam et al also reported that the impression that the alert system prioritized sensitivity over specificity limited perceived credibility for some users. Our evaluation also demonstrated the presence of both uncertainties and variations in professional judgement among specialists, as to what changes in serum creatinine were clinically significant. In addition, respondents in our evaluation were occasionally unsure as to whether clinical intervention was warranted, even for cases where the AKI alert was perceived to be genuine. Bevan's [36] exploration of the impact of a clinical decision support system for AKI embedded within a single hospital's EPR found that alerts were unpopular because of their interruption to the established workflows. We were able to avoid this problem through the separation of alerts from the hospital's EPR so that mobile working allowed users to integrate alert reviews into their routine working practice. Our finding that mobile working tools are integrated into clinicians' workflow is pertinent given that junior medical staff currently spend almost half their working day using desktop computers [37]. Finally, the survey by Oh et al [38] of provider acceptance of automated electronic alerts for AKI demonstrated that approval of alerts was positively correlated with the belief that such alerts improved patient care and negatively correlated with the belief that alerts did not provide any novel information. The overall odds of approval decreased over time. Thus, the success of deploying clinical alerts is dependent on clinicians' perceptions of their relevance. This tallies with our finding of diverse opinions about the value of low risk alerts. A number of mixed-methods analyses of electronic alerting systems for AKI are still underway; results from the qualitative segments of the *Acute Kidney Outreach to Reduce Deterioration and Death* [39] and *Tackling AKI* [40] studies are awaited.

### Conclusions

Our results are relevant to the design and evaluation of care pathways that involve automated alerts, mobile working, or the early deployment of specialist care. Such innovations will increasingly emerge with the application of machine learning [41,42] to early diagnosis and disease prediction. Our findings suggest that alerting systems aiming to encourage early or preventive action will achieve buy-in from health care professionals if they believe in the clinical and cost-effectiveness of the intervention, feel equipped to respond, understand clearly what their responsibilities are, and feel empowered to act. Training in prioritization of information is needed to balance the planned benefits of real-time access to mobile data with the cognitive load this will produce; digitally enabled pathways should be designed so that the most appropriate clinician is able to access the right data at the right time. E-alerting or the early deployment of a specialist resource may also have an impact on other clinical teams affected by implementation; future evaluations should seek to further explore this. Finally, the inevitable introduction of digital technology to health care is more likely to improve both patient outcomes and working practices if aligned with a commitment to proactively identify and address concomitant, and sometimes unexpected, sequelae.


## Acknowledgments

RR and GB are in part supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care North Thames at Barts Health NHS Trust, and RR is an NIHR senior investigator. GR and HM are funded in part by the NIHR University College London Hospitals Biomedical Research Centre. OSA is in part supported by an NIHR academic clinical fellowship. The authors wish to thank the staff and patients of the RFH and the RFH data monitoring committee. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.


## Authors' Contributions

HM, CL, RR, CH, AK, TB, KA, DK, and MS initiated the project and the collaboration. CL conceived the care pathway. AC, CL, CM, JC, GJ, SS, and ME supported implementation. RR led the design of the evaluation with assistance from AC, CL, GR, HM, PM, and CN. AC and CL triaged alerts necessary for the evaluation. AC collected all necessary data that were analyzed by AC with assistance and oversight from GB and RR. AC, HM, RR, PM, CL, OSA, GB, and GR wrote the paper. All authors read and agreed to the final submission.

## Conflicts of Interest

CL, HM, GR, and RR are paid clinical advisors to DeepMind. AC's clinical research fellowship was part-funded by DeepMind. CL was a member of the National Institute for Health and Care Excellence clinical guideline 169 development group referenced in the paper. DeepMind remained independent from the collection and analysis of all data. HM coholds a patent on a fluid delivery device that might ultimately help in preventing some (dehydration-related) cases of AKI occurring. DeepMind was acquired by Google in 2014 and is now a part of the Alphabet group. The deployment of Streams at RFH was the subject of an investigation by the Information Commissioner's Office in 2017. RFH has since published an audit completed to comply with undertakings following this investigation [43]. In November 2018, it was announced that the Streams team will be joining Google as part of a wider health effort [44].

## Multimedia Appendix 1

Care Protocol, nursing advisory sticker, interview guide, and Royal Free Hospital committee members.

[PDF File (Adobe PDF File), 402KB - jmir_v21i7e13143_app1.pdf ]

## Multimedia Appendix 2

Matrix of quotes from interviews. Rows represent individual respondents, and columns represent categories which align to the basic principles of the intervention pathway.

[XLSX File (Microsoft Excel File), 44KB - jmir_v21i7e13143_app2.xlsx ]

## References

1. Weller J, Boyd M, Cumin D. Teams, tribes and patient safety: overcoming barriers to effective teamwork in healthcare. Postgrad Med J 2014 Mar;90(1061):149-154. [doi: 10.1136/postgradmedj-2012-131168] [Medline: 24398594]
2. Schwendimann R, Blatter C, Dhaini S, Simon M, Ausserhofer D. The occurrence, types, consequences and preventability of in-hospital adverse events - a scoping review. BMC Health Serv Res 2018 Dec 4;18(1):521 [FREE Full text] [doi: 10.1186/s12913-018-3335-z] [Medline: 29973258]
3. Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007 Apr;33(4):667-679. [doi: 10.1007/s00134-007-0532-3] [Medline: 17318499]
4. AHRQ Patient Safety Network. 2019. Rapid Response Systems URL: https://psnet.ahrq.gov/primers/primer/4 [accessed 2018-11-06]
5. Winters BD, Weaver SJ, Pfoh ER, Yang T, Pham JC, Dy SM. Rapid-response systems as a patient safety strategy: a systematic review. Ann Intern Med 2013 Mar 5;158(5 Pt 2):417-425 [FREE Full text] [doi: 10.7326/0003-4819-158-5-201303051-00009] [Medline: 23460099]
6. Smith F. Who invented that bleeping thing? Br Med J 2003 Sep 27;327(7417):719 [FREE Full text] [doi: 10.1136/bmj.327.7417.719]
7. Gov.uk. 2018. The Future of Healthcare: Our Vision for Digital, Data and Technology in Health and Care URL: https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care [accessed 2018-11-12]
8. Gottlieb S. US Food & Drug Administration. 2018. Transforming FDA's Approach to Digital Health URL: https://www.fda.gov/newsevents/speeches/ucm605697.htm [accessed 2018-11-14]
9. KDIGO Working Group. KDIGO. 2012. KDIGO clinical practice guidelines for acute kidney injury URL: https://kdigo.org/wp-content/uploads/2016/10/KDIGO-2012-AKI-Guideline-English.pdf
10. Porter CJ, Juurlink I, Bisset LH, Bavakunji R, Mehta RL, Devonald MA. A real-time electronic alert to improve detection of acute kidney injury in a large teaching hospital. Nephrol Dial Transplant 2014 Oct;29(10):1888-1893. [doi: 10.1093/ndt/gfu082] [Medline: 24744280]
11. Kerr M, Bedford M, Matthews B, O'Donoghue D. The economic impact of acute kidney injury in England. Nephrol Dial Transplant 2014 Jul;29(7):1362-1368. [doi: 10.1093/ndt/gfu016] [Medline: 24753459]
12. der Mesropian PJ, Kalamaras JS, Eisele G, Phelps KR, Asif A, Mathew RO. Long-term outcomes of community-acquired versus hospital-acquired acute kidney injury: a retrospective analysis. Clin Nephrol 2014 Mar;81(3):174-184. [doi: 10.5414/CN108153] [Medline: 24361059]
13. Aitken E, Carruthers C, Gall L, Kerr L, Geddes C, Kingsmore D. Acute kidney injury: outcomes and quality of care. QJM 2013 Apr;106(4):323-332. [doi: 10.1093/qjmed/hcs237] [Medline: 23345468]
14. Wang HE, Muntner P, Chertow GM, Warnock DG. Acute kidney injury and mortality in hospitalized patients. Am J Nephrol 2012;35(4):349-355 [FREE Full text] [doi: 10.1159/000337487] [Medline: 22473149]
15. Chawla LS, Amdur RL, Amodeo S, Kimmel PL, Palant CE. The severity of acute kidney injury predicts progression to chronic kidney disease. Kidney Int 2011 Jun;79(12):1361-1369 [FREE Full text] [doi: 10.1038/ki.2011.42] [Medline: 21430640]
16. Alleway R. National Confidential Enquiry Into Patient Outcome and Death. 2009. Acute Kidney Injury: Adding Insult to Injury (2009) URL: http://www.ncepod.org.uk/2009aki.html [accessed 2016-03-08]
17. Selby NM, Hill R, Fluck RJ, NHS England 'Think Kidneys' AKI Programme. Standardizing the early identification of acute kidney injury: the NHS England national patient safety alert. Nephron 2015;131(2):113-117 [FREE Full text] [doi: 10.1159/000439146] [Medline: 26351847]
18. Hill R, Selby N. Think Kidneys. 2014. Acute Kidney Injury Warning Algorithm: Best Practice Guidance URL: https://www.thinkkidneys.nhs.uk/wp-content/uploads/2014/12/AKI-Warning-Algorithm-Best-Practice-Guidance-final-publication-0112141.pdf [accessed 2018-11-14]
19. Wilson FP, Shashaty M, Testani J, Aqeel I, Borovskiy Y, Ellenberg SS, et al. Automated, electronic alerts for acute kidney injury: a single-blind, parallel-group, randomised controlled trial. Lancet 2015 May 16;385(9981):1966-1974 [FREE Full text] [doi: 10.1016/S0140-6736(15)60266-5] [Medline: 25726515]
20. Connell A, Montgomery H, Morris S, Nightingale C, Stanley S, Emerson M, et al. Service evaluation of the implementation of a digitally-enabled care pathway for the recognition and management of acute kidney injury. F1000Res 2017;6:1033 [FREE Full text] [doi: 10.12688/f1000research.11637.2] [Medline: 28751970]
21. Connell A, Montgomery H, Martin P, Nightingale C, Sadeghi-Alavijeh O, King D, et al. Evaluation of a digitally-enabled care pathway for the management of acute kidney injury in patients admitted to hospital as an emergency (in press). 2019.
22. Connell A, Montgomery H, Martin P, Nightingale C, Sadeghi-Alavijeh O, King D, et al. Implementation of a Digitally Enabled Care Pathway (Part 1): Impact on Clinical Outcomes and Associated Health Care Costs. J Med Internet Res 2019 Jul 31;21(7):e13147. [doi: 10.2196/13147]
23. DeepMind Health. How To Use Streams URL: https://support.deepmindhealth.com [accessed 2018-10-18]
24. London AKI Network. AKI Guidelines URL: http://www.londonaki.net/clinical/guidelines-pathways.html [accessed 2018-07-23]

25. National Institute for Health and Care Excellence. 2013. Acute Kidney Injury: Prevention, Detection and Management: Guidance URL: https://www.nice.org.uk/guidance/cg169 [accessed 2018-11-14]
26. Marshall MN. The key informant technique. Fam Pract 1996 Feb;13(1):92-97. [doi: 10.1093/fampra/13.1.92] [Medline: 8671109]
27. Marshall B, Cardon P, Poddar A, Fontenot R. Does sample size matter in qualitative research?: a review of qualitative interviews in is research. J Comput Inform Syst 2015 Dec 10;54(1):11-22. [doi: 10.1080/08874417.2013.11645667]
28. Hudelson PM. World Health Organization. 1996. Qualitative Research for Health Programmes URL: https://apps.who.int/iris/handle/10665/62315 [accessed 2018-07-13]
29. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. BMC Med Res Methodol 2018 Dec 21;18(1):148 [FREE Full text] [doi: 10.1186/s12874-018-0594-7] [Medline: 30463515]
30. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2006;5(1):80-92. [doi: 10.1177/160940690600500107]
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. NHS: Health Research Authority. Is My Study Research? URL: http://www.hra-decisiontools.org.uk/research/redirect.html [accessed 2018-03-23]
33. Sawhney S, Marks A, Ali T, Clark L, Fluck N, Prescott GJ, et al. Maximising acute kidney injury alerts--a cross-sectional comparison with the clinical diagnosis. PLoS One 2015;10(6):e0131909 [FREE Full text] [doi: 10.1371/journal.pone.0131909] [Medline: 26125553]
34. Alavijeh OS, Bansal J, Hadfield K, Laing C, Dawnay A. Implementation of an automated primary care acute kidney injury warning system: a quantitative and qualitative review of 2 years of experience. Nephron 2017;135(3):189-195. [doi: 10.1159/000452928] [Medline: 28030868]
35. Kanagasundaram NS, Bevan MT, Sims AJ, Heed A, Price DA, Sheerin NS. Computerized clinical decision support for the early recognition and management of acute kidney injury: a qualitative evaluation of end-user experience. Clin Kidney J 2016 Feb;9(1):57-62 [FREE Full text] [doi: 10.1093/ckj/sfv130] [Medline: 26798462]
36. Bevan M, Heed A, Sheerin NS, Sims A, Price DA, Kanagasundaram NS. Electronic clinical decision support for the early recognition and management of acute kidney injury: qualitative evaluation of end-user experience. Nephrol Dial Transplant 2015;30(Suppl 3):iii443. [doi: 10.1093/ndt/gfv190.10]
37. Méan M, Garnier A, Wenger N, Castioni J, Waeber G, Marques-Vidal P. Computer usage and task-switching during resident's working day: disruptive or not? PLoS One 2017;12(2):e0172878 [FREE Full text] [doi: 10.1371/journal.pone.0172878] [Medline: 28235078]
38. Oh J, Bia JR, Ubaid-Ullah M, Testani JM, Wilson FP. Provider acceptance of an automated electronic alert for acute kidney injury. Clin Kidney J 2016 Aug;9(4):567-571 [FREE Full text] [doi: 10.1093/ckj/sfw054] [Medline: 27478598]
39. Abdelaziz TS, Lindenmeyer A, Baharani J, Mistry H, Sitch A, Temple RM, et al. Acute kidney outreach to reduce deterioration and death (AKORDD) trial: the protocol for a large pilot study. BMJ Open 2016 Dec 19;6(8):e012253 [FREE Full text] [doi: 10.1136/bmjopen-2016-012253] [Medline: 27543592]
40. Selby NM, Casula A, Lamming L, Mohammed M, Caskey F, Tackling AKI Investigators. Design and rationale of 'tackling acute kidney injury', a multicentre quality improvement study. Nephron 2016;134(3):200-204 [FREE Full text] [doi: 10.1159/000447675] [Medline: 27376867]
41. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. Crit Care Med 2018 Jul;46(7):1070-1077. [doi: 10.1097/CCM.0000000000003123] [Medline: 29596073]
42. Mohamadlou H, Lynn-Palevsky A, Barton C, Chettipally U, Shieh L, Calvert J, et al. Prediction of acute kidney injury with a machine learning algorithm using electronic health record data. Can J Kidney Health Dis 2018;5:2054358118776326 [FREE Full text] [doi: 10.1177/2054358118776326] [Medline: 30094049]
43. Royal Free Hospital. 2018. Royal Free London Publishes Audit Into Streams App URL: https://www.royalfree.nhs.uk/news-media/news/royal-free-london-publishes-audit-into-streams-app/ [accessed 2018-10-18]
44. DeepMind Technologies. Scaling Streams with Google URL: https://deepmind.com/blog/scaling-streams-google/ [accessed 2018-11-21]

## Abbreviations

**AKI:** acute kidney injury
**EPR:** electronic patient record
**NHS:** National Health Service
**NIHR:** National Institute for Health Research
**PARRT:** patient-at-risk and resuscitation team
**RFH:** Royal Free Hospital

*Edited by G Eysenbach; submitted 14.12.18; peer-reviewed by M Devonald, J Varghese, N Selby; comments to author 17.01.19; revised version received 29.01.19; accepted 24.03.19; published 15.07.19.*

*Please cite as:*
*Connell A, Black G, Montgomery H, Martin P, Nightingale C, King D, Karthikesalingam A, Hughes C, Back T, Ayoub K, Suleyman M, Jones G, Cross J, Stanley S, Emerson M, Merrick C, Rees G, Laing C, Raine R*
*Implementation of a Digitally Enabled Care Pathway (Part 2): Qualitative Analysis of Experiences of Health Care Professionals*
*J Med Internet Res 2019;21(7):e13143*
*URL: http://www.jmir.org/2019/7/e13143/*
*doi:10.2196/13143*
*PMID:31368443*

Original Paper

# Implementation of a Digitally Enabled Care Pathway (Part 1): Impact on Clinical Outcomes and Associated Health Care Costs

Alistair Connell[1,2], MBBS; Rosalind Raine[3], PhD; Peter Martin[3], PhD; Estela Capelas Barbosa[3], PhD; Stephen Morris[3], PhD; Claire Nightingale[3,4], PhD; Omid Sadeghi-Alavijeh[5], MSc; Dominic King[2], PhD; Alan Karthikesalingam[2], PhD; Cían Hughes[2], MSc; Trevor Back[2], PhD; Kareem Ayoub[2], MA; Mustafa Suleyman[2]; Gareth Jones[5], PhD; Jennifer Cross[5], PhD; Sarah Stanley[5], MSc; Mary Emerson[5], BSc; Charles Merrick[5]; Geraint Rees[6], PhD; Hugh Montgomery[1*], MD; Christopher Laing[5*], MD

[1]Centre for Human Health and Performance, University College London, London, United Kingdom
[2]DeepMind Health, London, United Kingdom
[3]Department of Applied Health Research, University College London, London, United Kingdom
[4]Population Health Research Institute, St George’s, University of London, London, United Kingdom
[5]Royal Free London NHS Foundation Trust, London, United Kingdom
[6]Faculty of Life Sciences, University College London, London, United Kingdom
[*]these authors contributed equally

**Corresponding Author:**
Christopher Laing, MD
Royal Free London NHS Foundation Trust
Pond Street
London,
United Kingdom
Phone: 44 020 7794 0500 ext 33322
Email: chris.laing@nhs.net

## Abstract

**Background:** The development of acute kidney injury (AKI) in hospitalized patients is associated with adverse outcomes and increased health care costs. Simple *automated e-alerts* indicating its presence do not appear to improve outcomes, perhaps because of a lack of explicitly defined integration with a clinical response.

**Objective:** We sought to test this hypothesis by evaluating the impact of a digitally enabled intervention on clinical outcomes and health care costs associated with AKI in hospitalized patients.

**Methods:** We developed a care pathway comprising automated AKI detection, mobile clinician notification, in-app triage, and a protocolized specialist clinical response. We evaluated its impact by comparing data from pre- and postimplementation phases (May 2016 to January 2017 and May to September 2017, respectively) at the intervention site and another site not receiving the intervention. Clinical outcomes were analyzed using segmented regression analysis. The primary outcome was recovery of renal function to ≤120% of baseline by hospital discharge. Secondary clinical outcomes were mortality within 30 days of alert, progression of AKI stage, transfer to renal/intensive care units, hospital re-admission within 30 days of discharge, dependence on renal replacement therapy 30 days after discharge, and hospital-wide cardiac arrest rate. Time taken for specialist review of AKI alerts was measured. Impact on health care costs as defined by Patient-Level Information and Costing System data was evaluated using difference-in-differences (DID) analysis.

**Results:** The median time to AKI alert review by a specialist was 14.0 min (interquartile range 1.0-60.0 min). There was no impact on the primary outcome (estimated odds ratio [OR] 1.00, 95% CI 0.58-1.71; *P*=.99). Although the hospital-wide cardiac arrest rate fell significantly at the intervention site (OR 0.55, 95% CI 0.38-0.76; *P*<.001), DID analysis with the comparator site was not significant (OR 1.13, 95% CI 0.63-1.99; *P*=.69). There was no impact on other secondary clinical outcomes. Mean health care costs per patient were reduced by £2123 (95% CI −£4024 to −£222; *P*=.03), not including costs of providing the technology.

**Conclusions:** The digitally enabled clinical intervention to detect and treat AKI in hospitalized patients reduced health care costs and possibly reduced cardiac arrest rates. Its impact on other clinical outcomes and identification of the *active* components of the pathway requires clarification through evaluation across multiple sites.

***(J Med Internet Res 2019;21(7):e13147)*** doi:10.2196/13147

**KEYWORDS**

nephrology; acute kidney injury

## Introduction

### Background

Acute kidney injury (AKI)—a sudden decline in kidney function—can be caused by hypovolemia, infection (including severe sepsis), nephrotoxicity, primary renal diseases, and urinary tract obstruction [1]. Affecting more than 18% of hospitalized patients [2], it is associated with prolonged hospital stay, need for acute renal replacement therapy (RRT), or intensive care admission as well as the development of chronic kidney disease and the need for long-term dialysis [3-5]. Although AKI may be a marker of systemic physiological decompensation in acute illnesses (eg, sepsis, trauma, or high-risk surgery), AKI itself might directly cause additional deaths through, for instance, metabolic derangement or extracellular fluid volume overload [6]. Such impacts are expensive; AKI confers excess annual costs of £1 billion to the English National Health Service (NHS) [7]. Similar excess health care costs have been demonstrated in other health systems [8].

AKI management involves the identification and treatment of life-threatening complications and medical or surgical treatment of underlying cause, supportive care (including RRT, where necessary), and interventions to reduce risk of recurrence [9]. In response to poor outcomes and variations in care delivery [10], automated AKI alerts (using standardized definitions of its presence and severity based on increases in serum creatinine [11]) have been delivered using messages in electronic health record systems or through hospital pagers [12]. In England, this approach has been applied through the embedding of an AKI detection algorithm—The NHS Early Detection Algorithm, NHSEDA (Multimedia Appendix 1)—in laboratory information management systems [13]. However, evidence of the impact of electronic alerts in improving clinical outcomes in AKI is conflicting [14]. The greatest indications of improvement seem to occur when such detection systems are coupled with structured education and clinical intervention packages [15,16]. However, the delivery of such care pathways is challenging: AKI is common and of heterogeneous etiology; it presents in diverse settings; and it is normally, at least in its early stages, managed by a range of nonspecialist teams.

### Objectives

To address these issues, we developed a digitally enabled care pathway for AKI patients [17]. This uses a mobile app (Streams), which alerts a specialist response team to the presence of AKI in real time, simultaneously providing relevant clinical data in a user-friendly format and allowing communication of key triage decisions among team members. Members of the response team review patients using a care protocol that maps to best practice guidance [9].

We have reported the clinical impact of this digitally enabled care pathway on patients with AKI at the point of presentation to the emergency department (ED) [18]. The limited impacts we identified in this setting might reflect the difficulty of mitigating harm when AKI is well established or in the context of pathogeneses specific to community-acquired AKI. In this paper, we assess the impact of the care pathway on clinical outcomes for patients who develop AKI following hospital admission and on health care costs.

## Methods

### The Hospital Sites

The digital pathway was implemented at the Royal Free Hospital (RFH), a large (839 beds including a 34-bed intensive treatment unit [ITU]) hospital in north London, United Kingdom. It provides acute and emergency care as well as a range of specialist, regional inpatient services (eg, hepatology, HIV and infectious disease, amyloidosis, and vascular surgery) and has a large inpatient nephrology and renal transplant service.

For the purposes of our evaluation, we used a comparator site managed by the same health care provider organization (Royal Free London NHS Foundation Trust, RFLFT) in which the intervention was not implemented. Barnet General Hospital (BGH) is an acute general hospital with 459 beds. It has a 21-bed ITU that can provide acute RRT and a liaison nephrology service. Tertiary, specialist services are not provided on this site. A number of parallel improvement initiatives were ongoing at the comparator site during the study period, including a sepsis improvement project and an active deteriorating patients improvement program.

### Implementation

Blood tests, including serum creatinine, are routinely undertaken on hospitalized inpatients across all wards as directed by the treating clinicians. Historically, at both sites, blood tests would be reviewed in batches by the clinicians who ordered them. Results suggesting AKI would be telephoned to relevant wards by laboratory staff. Referral for nephrology assessment would be undertaken at the discretion of the clinical teams and using hospital pagers and phones. Cases would be prioritized and treated by the nephrology teams through assessment of referral information and results on desktop computers and through bedside review. The Patient at Risk and Resuscitation Team (PARRT) provides support to ward teams for patients deemed at risk of deterioration or who trigger existing, physiology-based early warning systems.

The digitally enabled AKI care pathway and the technical architecture of the Streams app have been described in detail previously [17], and the pre-existing and novel care pathways are shown in Multimedia Appendix 1. Members of the response team undertook training before implementation. Following implementation, Streams continuously applied the NHSEDA to creatinine results for all inpatients. Using iPhones (Apple Inc), the nephrology team was alerted to all potential cases of

AKI, with their AKI stage and whether metabolic complications such as hyperkalemia were present. A curated dataset was provided, which included patients' demographic characteristics, previously coded diagnoses, and relevant results. Filters excluded children, critical care and chronic dialysis patients, and those already under the care of the nephrology inpatient team from producing alerts. Cases were triaged in-app and, when clinical review was warranted, a best-practice care protocol was delivered (Multimedia Appendix 1). This was annotated and entered into the patient's notes alongside an advisory sticker for key nursing actions (Multimedia Appendix 1). Recovery could be monitored in-app, and repeat AKI alerts were sent if AKI had not recovered after 48 hours or if AKI severity stage increased. Nephrology members received all AKI alerts, and PARRT members received alerts only for patients with stages 2 and 3 AKI. All team members could communicate in-app; triage responses and the outcome of clinical reviews were visible in the app to other team members. Implementation of the care pathway at the RFH used existing RFH PARRT and nephrology staff and did not result in expansion in staff numbers. A diagram outlining the pre- and postintervention care pathways is provided in Multimedia Appendix 1.

### Data Collection

At both sites, data from RFLFT hospital databases and those supporting Streams app relating to the intervention period (May to September 2017) were compared with those from a predeployment phase (May 2016 to January 2017). Data relating to patients in whom an AKI alert was generated on presentation to the hospital ED are reported elsewhere [18], and such patients were excluded from the analyses reported here.

Data collected and their sources are detailed in Table 1. The time frame to alert viewing was determined using data recorded by the Streams app. The presence of individual comorbidities and overall patient-specific Charlson comorbidity index score (which categorizes comorbidities based on the International Classification of Diseases diagnosis codes) were derived as per the method by Thygesen et al [19]. Patients were sorted into national quintiles of deprivation (quintile 1, least; quintile 5, most deprived) using Indices of Multiple Deprivation (IMD)—a measure combining 7 domains (income, employment, living environment, health, education skills and training deprivation and disability, barriers to housing and services, and crime) into a single deprivation score for a small area—by cross-referencing patients' postcodes with the UK Government's Indices of Deprivation 2015 dataset [20].

For the economic analysis, we used Payment Level Information and Costing System (PLICS) data supplied by the RFLFT. PLICS is a clinical costing system where costs are derived for each patient spell (ie, admission) by tracing resources used by an individual patient in diagnosis and treatment and calculating the expenditure on those resources using the actual costs incurred by the provider. PLICS has the advantage of including staffing costs and infrastructure absorbed costs. In our study, the PLICS data for hospitalized patients with AKI included the following components: total length of stay (including the length of stay in intensive care unit), pathology and radiology examinations, total theater time, theater cutting time, inpatient dialysis, and overhead costs. These data were analyzed at the spell level. We also obtained data on the costs associated with selected individual components of a spell, which we analyzed separately (ie, length of stay, pathology and radiology examinations, theater total time, and theater cutting time). However, individual cost components were based on tariffs and not fully absorbed costs. Furthermore, we could not obtain individual costs of inpatient dialysis. The final dataset used in the economic analysis comprised total and component-specific spell-level costs at the RFH and BGH, before and after the digitally enabled care pathway was introduced at the RFH.

### Evaluation of Impacts

The primary outcome was recovery of renal function (return to a serum creatinine concentration within 120% of the baseline, as defined by the NHSEDA) before hospital discharge. Table 1 describes the predefined secondary endpoints. At both sites, NHSEDA was used to identify potential AKI cases. Because the NHSEDA can produce false positives [22], 2 authors (AC and CL) clinically validated all AKI alerts produced from all periods at both hospital sites. Only clinician-confirmed episodes of AKI were included in the analysis. In this paper, we report the outcomes of inpatients producing AKI alerts outside of the ED during the predeployment and deployment phases (Figure 1). The impact of the care pathway on cardiac arrests rate was measured on a hospital level, as it was not possible to ascertain which cardiac arrests occurred among patients with AKI.

**Table 1.** Definitions of each outcome and sources of data collected.

| Data category and measure | Definition | Source of data |
|---|---|---|
| **Sociodemographic characteristics** | | |
| Age | Age in years at the time of alert | HL7[a] data aggregated within the Streams data processor |
| Gender | Gender codes used in the NHS[b] Data Dictionary [21] | HL7 data aggregated within the Streams data processor |
| Ethnicity | Ethnicity category codes used in the NHS Data Dictionary [21] | HL7 data aggregated within the Streams data processor |
| Comorbid disease | Presence of individual Charlson index comorbidities and overall Charlson score | HL7 data aggregated within the Streams data processor |
| Deprivation | Index of Multiple Deprivation | Ministry of Housing, Communities and Local Government database |
| **Clinical outcomes** | | |
| Recovery of renal function | Return to <120% index creatinine (as defined by NHSEDA[c]) by the time of hospital discharge | HL7 data aggregated within the Streams data processor |
| Time to recovery of renal function | The time from AKI[d] alert to recovery of renal function (<120% index creatinine) | HL7 data aggregated within the Streams data processor |
| Mortality | Death in 30 days following AKI alert | HL7 data aggregated within the Streams data processor |
| Progression of AKI stage | Movement between AKI severity classes following AKI alert and before hospital discharge | HL7 data aggregated within the Streams data processor |
| Admission to high acuity or specialist renal inpatient bed | Admission to acute kidney unit/high dependency unit/intensive treatment unit during index admission | HL7 data aggregated within the Streams data processor |
| Requirement long-term renal replacement therapy | Use of hemofiltration/hemodiafiltration/hemodialysis/peritoneal dialysis in 30 days following hospital discharge date | RFH[e] Nephrology Clinical Information Management System |
| Length of stay | Time from AKI alert to hospital discharge | HL7 data aggregated within the Streams data processor |
| Re-admission to hospital | Re-admission to hospital in 30 days following index admission discharge date | HL7 data aggregated within the Streams data processor |
| **Trust-wide metric** | | |
| Cardiac arrest rate | Number of cardiac arrests per 1000 bed days | Trust critical care nursing team logs |
| **Economic measures** | | |
| Costs per patient | Cost per patient per hospital spell | Payment Level Information and Costing System data and Payment by Results/local tariffs at the trust |
| **Process of care** | | |
| Time to alert review | Time from alert generation to alert viewing by a clinician | Data aggregated within the Streams data processor |

[a]Health Level 7 (HL7) messages are used to transfer information between different health care information technology systems.

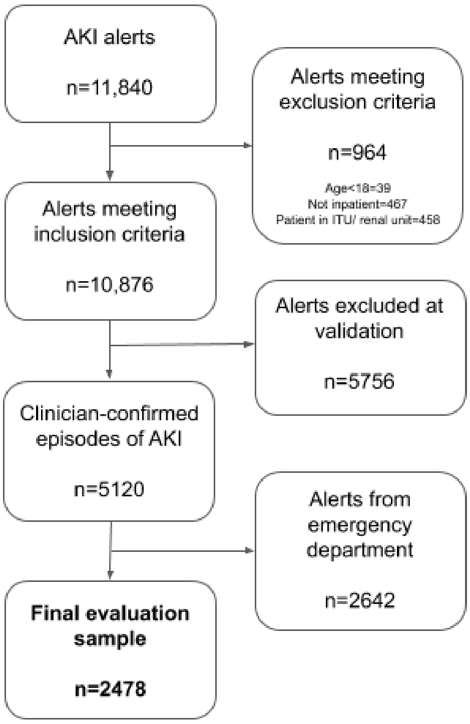
[b]NHS: National Health Service.

[c]NHSEDA: NHS Early Detection Algorithm.

[d]AKI: acute kidney injury.

[e]RFH: Royal Free Hospital.

**Figure 1.** Defining the final evaluation sample. AKI: acute kidney injury; ITU: intensive treatment unit.




## Statistical Analysis

All data were pseudonymized before transfer to the University College London (UCL) for analysis. Analyses were performed using R, version 3.4.3 (R core team) [23], and Stata MP version 14 (StataCorp) [24]. Segmented regression analysis estimated the intervention effect on the primary outcome (return to a serum creatinine concentration within 120% of the baseline, as defined by the NHSEDA) and 5 secondary outcome measures: mortality within 30 days of alert, progression of AKI stage, transfer to renal/intensive care units during admission, re-admission within 30 days of discharge, and dependence on RRT 30 days after discharge. Outcomes were measured as weekly proportions. We used binomial regression models with a logit link predicting the weekly rate of each outcome. Codes 1 and 0 were applied to the period after and before the intervention, respectively. The intervention and comparator sites were coded 1 and 0, respectively. The variable *time* denoted the week number (1 denoting the first week of the intervention period, and negative numbers denoting weeks in the preintervention period). The statistical model used was:

$$\text{logit}(y) = \beta_0 + \beta_1\text{int} + \beta_2\text{time} + \beta_3\text{site} + \beta_4\text{int x time} + \beta_5\text{int x site} + \beta_6\text{time x site } \beta_7\text{int x time x site} \quad (1)$$

where the proportion of interest is denoted by *y*, the variables intervention, time, and site by *int*, *time*, and *site*, respectively (as defined previously), and the coefficients to be estimated by $\beta_0$,...,. In addition, 2 coefficients evaluated the evidence for the intervention causing a step change in outcome: the effect of *intervention* estimates the step change in outcome at the start of the intervention period at the RFH. The interaction *site*×*intervention* estimates the difference-in-differences (DID) in the step change between the intervention and comparator hospital sites. We also evaluated evidence for a change in temporal trend in the outcome because of the intervention: the *time*×*intervention* interaction estimates the difference in temporal outcome trend between the intervention and preintervention periods at the RFL; the 3-way *time*×*site*×*intervention* interaction estimates the DID in the trend between the intervention and comparator sites.

For all models, we inspected the autocorrelation function (up to lag 15). No significant autocorrelation was detected in any model. At the point of protocol publication, it was not anticipated that we would be able to collect patient-level data relating to sociodemographic characteristics and comorbid disease.

To examine the robustness of our primary outcome analysis, we used binary logistic regression to perform a sensitivity analysis: the same model mentioned previously was used, except that (1) the outcome was defined at the patient level and (2) patient-level characteristics (age, sex, ethnicity category, index of multiple deprivation, AKI alert level, the presence of complications at the time of alert, and the presence of individual Charlson score comorbidities) were included as covariates to adjust for any differences in casemix between sites and within sites over time.

The Wilcoxon rank-sum test was used to analyze the time to creatinine recovery (where this occurred by hospital discharge). To allow for the effects of in-hospital death on this outcome, the effect of the intervention on the length of hospital stay was estimated by competing risk analysis [25]. To determine the effect of the intervention on the time to recognition of AKI, a survival analysis was performed. The Wilcoxon rank-sum and chi-square tests were used to analyze sociodemographic variables as appropriate.

A total of 500 alerts were selected randomly from all periods, and all sites were reviewed a second time to assess the reliability of case validation. Intra- and interrater reliability was determined using Cohen's kappa coefficient (Multimedia Appendix 1).

The number of cardiac arrests was recorded monthly at both hospital sites. Data relating to those which occurred in the hospitals' ED, cardiac catheterization laboratory, intensive care unit, coronary care unit or in patients who had a formal *not for resuscitation* order signed were not included in the monthly counts recorded at the hospital level. Poisson regression models with a log link and an offset variable adjusting for the number of admissions per month were used to estimate the intervention effect on this outcome. As data were collected monthly, there was a relative paucity of postintervention data points so that estimating the effect of the intervention on outcome trend was not possible. The statistical model was:

$$\log(\text{number of cardiac arrests}) = \beta_0 + \beta_1\text{int} + \beta_2\text{site} + \beta_3\text{int x site} + \log(\text{number of admissions}) \quad (2)$$

Economic analyses used generalized linear models (GLMs) to estimate DID, where costs were defined at the spell level, and patient-level characteristics (age, sex, ethnicity category, IMD, the presence of complications at the time of alert, and the presence of individual Charlson score comorbidities, such as diabetes mellitus or congestive cardiac failure) were included as covariates so as to allow adjustment for any differences in casemix between sites and within sites over time. A GLM was specified using a gamma family and log link to account for data skewness. The model used was:

$$\log(\text{cost}) = \beta_0 + \beta_1\text{age} + \beta_2\text{sex} + \beta_3\text{ethnicity} + \beta_4\text{imd} + \beta_5\text{comp} + \beta_6\text{CharlsonScore} + \beta_7\text{time} + \beta_8\text{site} + \beta_9\text{time} \times \text{site} \quad (3)$$

where *time* was defined in relation to the intervention. May to September 2016 was considered preintervention ($t_1$), and May to September 2017 was considered postintervention ($t_3$). For robustness checks, we also carried out a secondary analysis, where the preintervention period was May 2016 to January 2017. The coefficient $\beta_9$ is the coefficient of interest, measuring the between-site DID, comparing the change over time at the RFH to the change over time at the BGH. We present predictive margins showing adjusted mean costs per spell at the RFH and BGH before and after the intervention was introduced at the RFH. We adjusted for clustering at the patient level to account for the possibility that patients may have had multiple spells.

### Ethical Approval

The digitally enabled care pathway constituted a new standard service at the RFH. The UCL Joint Research Office reviewed the study protocol and judged that the project fell under the remit of service evaluation as per guidance from the NHS Health Research Authority [26]. As such, no patient consent was required. The evaluation was registered with the RFH Audit Lead and Medical Director. An independent data monitoring committee (which included a patient member) reviewed all analyses before preparation for publication. A full list of committee members is provided in Multimedia Appendix 1.

## Results

Alerts produced for hospitalized patients during the intervention period were reviewed by a member of the specialist response team in a median time of 14.0 min (interquartile range [IQR] 1.0-60.0 min). At the intervention site, clinical validation of the 4392 and 2254 AKI alerts during predeployment (May 2016 to January 2017) and postdeployment (May to September 2017) phases, respectively, yielded 1760 and 919 inpatient AKI episodes in each phase. Of these, 56.5% (994/1960) and 52.2% (480/919), respectively, were located outside the ED. In the predeployment and postdeployment phases at the nonintervention site, clinical validation of the 2866 and 1364 alerts, respectively, yielded 1669 and 772 inpatient AKI episodes, with 39.2% (654/1669) and 45.3% (350/772) being located outside the ED.

Table 2 summarizes the sociodemographic and clinical characteristics of patients producing AKI alerts at both sites and periods. RFH inpatients were younger (median 72 vs 82 years, *P*<.001), less likely to be white (*P*<.001), and less deprived (*P*<.001) than at BGH. RFH patients had significantly less comorbidity (median [IQR] Charlson score 5.0 [3.0-8.0] vs 5.0 [4.0-8.0], *P*<.001). The proportion of patients with pre-existing renal disease was also lower (31.5% vs 37.2%, *P*<.001). Comparing the pre- and postintervention cohorts, there were some significant differences within the comparator site. At BGH, patients in the postintervention period had significantly more severe AKI (*P*=.01) and a higher burden of comorbid (*P*<.001) and renal disease (45.1% vs 32.9%, *P*<.001) than patients in the preintervention period.

**Table 2.** Sociodemographic and clinical characteristics of patients producing acute kidney injury alerts.

| Variable | Hospital site/period | | | | *P* value | | |
|---|---|---|---|---|---|---|---|
| | RFH[a] | | BGH[b] | | RFH pre vs RFH post | BGH pre vs BGH post | All RFH vs all BGH |
| | Pre[c] | Post[d] | Pre | Post | | | |
| AKI[e] alerts, n | 994 | 480 | 654 | 350 | —[f] | — | — |
| **Alert severity, n (%)** | | | | | .102 | .01 | .32 |
| AKI1 | 809 (81.4) | 411 (85.6) | 571 (87.3) | 281 (80.3) | | | |
| AKI2 | 127 (12.8) | 44 (9.2) | 60 (9.2) | 47 (13.4) | | | |
| AKI3 | 58 (5.8) | 25 (5.2) | 23 (3.5) | 22 (6.3) | | | |
| Male, n (%) | 541 (54.4) | 257 (53.5) | 331 (50.6) | 186 (53.1) | .74 | .48 | .30 |
| Age (years), median (interquartile range) | 73.00 (58.00-84.00) | 7.00 (57.00-83.00) | 82.00 (73.00-88.00) | 82.00 (73.25-88.75) | .14 | .81 | <.001 |
| **Ethnicity, n (%)** | | | | | .09 | .32 | <.001 |
| White | 625 (62.9) | 281 (58.5) | 512 (78.3) | 274 (78.3) | | | |
| Black or Black British | 76 (7.7) | 34 (7.1) | 29 (4.4) | 12 (3.4) | | | |
| Asian or Asian British | 110 (11.1) | 52 (10.8) | 60 (9.2) | 25 (7.1) | | | |
| Mixed | 10 (1.0) | 2 (0.42) | 3 (0.5) | 4 (1.1) | | | |
| Other ethnic groups | 173 (17.4) | 111 (23.1) | 50 (7.7) | 35 (10.0) | | | |
| **Index of Multiple Deprivation, n (%)** | | | | | .87 | .83 | <.001 |
| Quintile 1 (least deprived) | 184 (18.5) | 84 (17.5) | 42 (6.42) | 25 (7.1) | | | |
| Quintile 2 | 216 (21.7) | 130 (27.1) | 132 (20.2) | 60 (17.1) | | | |
| Quintile 3 | 233 (23.4) | 89 (18.5) | 183 (28.0) | 111 (31.7) | | | |
| Quintile 4 | 224 (22.5) | 111 (23.1) | 186 (28.4) | 99 (28.3) | | | |
| Quintile 5 (most deprived) | 97 (9.8) | 46 (9.6) | 108 (16.5) | 53 (15.1) | | | |
| Unknown | 40 (4.0) | 20 (4.2) | 3 (0.5) | 2 (0.6) | | | |
| **Charlson Score, n (%)** | | | | | .49 | <.001 | <.001 |
| 0 | 114 (11.5) | 49 (10.2) | 10 (1.5) | 7 (2.0) | | | |
| 1 | 51 (5.13) | 11 (2.3) | 25 (3.8) | 9 (2.6) | | | |
| 2 | 63 (6.3) | 54 (11.2) | 29 (4.4) | 13 (3.7) | | | |
| 3 | 107 (1.8) | 43 (9.0) | 78 (11.9) | 21 (6.0) | | | |
| 4 | 169 (17.0) | 63 (13.1) | 150 (22.9) | 59 (16.9) | | | |
| ≥5 | 490 (49.3) | 260 (54.2) | 362 (55.4) | 241 (68.9) | | | |
| Pre-existing renal disease present, n (%) | 303 (30.5) | 162 (33.8) | 215 (32.9) | 158 (45.1) | .23 | <.001 | <.001 |

[a]RFH: Royal Free Hospital.
[b]BGH: Barnet General Hospital.
[c]Pre: May 2016 to January 2017.
[d]Post: May 2017 to September 2017.
[e]AKI: acute kidney injury.
[f]Not applicable.

**Table 3.** Descriptive statistics of total cost per spell producing acute kidney injury alerts.

| Statistics | Royal Free Hospital (£) | | Barnet General Hospital (£) | |
|---|---|---|---|---|
| | Pre[a] | Post[b] | Pre | Post |
| Mean (SD) | 12,015.24 (22,732.78) | 10,154.92 (19,582.30) | 7391.16 (14,346.27) | 7108.88 (11,512.95) |
| Median | 5640.50 | 4954.00 | 3712.50 | 3774.00 |
| 1st centile | 166.00 | 207.00 | 160.00 | 199.00 |
| 25th centile | 2391.50 | 2079.00 | 1424.00 | 1153.50 |
| 75th centile | 13,208.50 | 10,567.00 | 8466.00 | 8897.00 |
| 99th centile | 111,245.00 | 90,138.00 | 51,991.00 | 45,614.00 |

[a]Pre: May 2016 to January 2017.

[b]Post: May 2017 to September 2017.

Table 3 provides descriptive statistics of total costs per spell at each site before and after the intervention. Multimedia Appendix 1 shows the positively skewed distribution of these costs.

## Clinical Outcomes

Estimates from the models predicting clinical outcomes are reported in Tables 4-7, as far as they relate to the research hypotheses. All estimated model coefficients are reported in Multimedia Appendix 1.

**Table 4.** Results of segmented regression analyses for renal recovery and mortality.

| Variable/interaction term | Renal recovery | | | Mortality | | |
|---|---|---|---|---|---|---|
| | Beta | *P* value | OR[a] (95% CI) | Beta | *P* value | OR (95% CI) |
| Intervention[b] | .00 | .99 | 1.00 (0.58-1.71) | .17 | .67 | 1.18 (0.55-2.52) |
| Site×intervention[c] | .22 | .62 | 1.24 (0.53-2.92) | .06 | .91 | 1.07 (0.36-3.15) |
| Time×intervention[d] | −.01 | .61 | 0.99 (0.96-1.03) | .00 | .89 | 1.00 (0.96-1.05) |
| Time×site×intervention[e] | −.03 | .29 | 0.97 (0.92-1.03) | −.03 | .44 | 0.97 (0.91-1.04) |

[a]OR: odds ratio.

[b]The coefficient *intervention* provides an estimate of the difference in outcome between the intervention period and the preintervention period at RFH.

[c]The 2-way interaction *site*×*intervention* provides an estimate of the difference-in-difference between the 2 hospital sites.

[d]The 2-way interaction *time*×*intervention* provides an estimate of the difference in outcome trend over time in the intervention period compared with the preintervention period at RFH.

[e]The 3-way interaction *time*×*site*×*intervention* provides an estimate of the difference-in-difference in the trend between the sites.

**Table 5.** Results of segmented regression analyses for progression of acute kidney injury stage and admission to intensive treatment unit/renal unit.

| Variable/interaction term | Progression of acute kidney injury stage | | | Admission to intensive treatment unit/renal unit | | |
|---|---|---|---|---|---|---|
| | Beta | *P* value | OR[a] (95% CI) | Beta | *P* value | OR (95% CI) |
| Intervention[b] | .67 | .11 | 1.96 (0.86-4.47) | .40 | .42 | 1.50 (0.57-4.00) |
| Site×intervention[c] | −.71 | .27 | 0.49 (0.14-1.71) | −1.18 | .18 | 0.31 (0.05-1.68) |
| Time×intervention[d] | −.01 | .60 | 0.99 (0.93-1.04) | .02 | .55 | 1.02 (0.96-1.08) |
| Time×site×intervention[e] | .04 | .32 | 1.04 (0.96-1.13) | .07 | .19 | 1.08 (0.97-1.20) |

[a]OR: odds ratio.

[b]The coefficient *intervention* provides an estimate of the difference in outcome between the intervention period and the preintervention period at RFH.

[c]The 2-way interaction *site*×*intervention* provides an estimate of the difference-in-difference between the 2 hospital sites.

[d]The 2-way interaction *time*×*intervention* provides an estimate of the difference in outcome trend over time in the intervention period compared with the preintervention period at RFH.

[e]The 3-way interaction *time*×*site*×*intervention* provides an estimate of the difference-in-difference in the trend between the sites.

**Table 6.** Results of segmented regression analyses for hospital re-admission and renal replacement therapy use.

| Variable/interaction term | Re-admission at 30 days | | | Renal replacement therapy use at 30 days | | |
|---|---|---|---|---|---|---|
| | Beta | *P* value | OR[a] (95% CI) | Beta | *P* value | OR (95% CI) |
| Intervention[b] | .20 | .54 | 1.22 (0.65-2.29) | −3.32 | .03 | 0.04 (0.00-0.62) |
| Site×intervention[c] | −.16 | .77 | 0.86 (0.31-2.39) | −1.04 | .99 | 0.35 (0-infinity) |
| Time×intervention[d] | −.03 | .23 | 0.97 (0.93-1.02) | .00 | .98 | 1.00 (0.83-1.23) |
| Time×site×intervention[e] | .01 | .84 | 1.01 (0.94-1.08) | −17.62 | .99 | 0.00 (0-infinity) |

[a]OR: odds ratio.

[b]The coefficient *intervention* provides an estimate of the difference in outcome between the intervention period and the preintervention period at RFH.

[c]The 2-way interaction *site×intervention* provides an estimate of the difference-in-difference between the 2 hospital sites.

[d]The 2-way interaction *time×intervention* provides an estimate of the difference in outcome trend over time in the intervention period compared with the preintervention period at RFH.

[e]The 3-way interaction *time×site×intervention* provides an estimate of the difference-in-difference in the trend between the sites.

**Table 7.** Results of segmented regression analysis for hospital cardiac arrest rate

| Variable/interaction term | Cardiac arrests | | |
|---|---|---|---|
| | Beta | *P* value | OR (95% CI) |
| Intervention[a] | −.60 | <.001 | 0.55 (0.38-0.76) |
| Site×intervention[b] | .12 | .69 | 1.13 (0.63-1.99) |

[a]The coefficient *intervention* provides an estimate of the difference in outcome between the intervention period and the preintervention period at RFH.

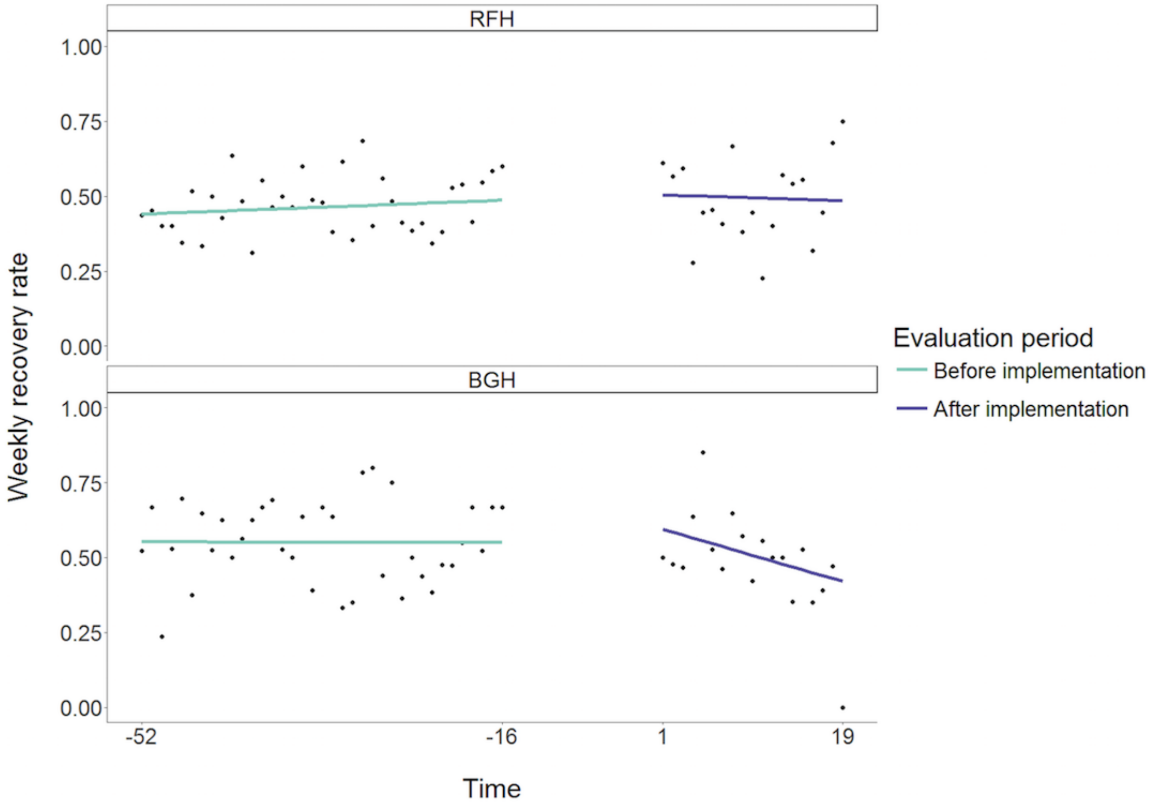
[b]The 2-way interaction *site×intervention* provides an estimate of the difference-in-difference between the 2 hospital sites.

### *Primary Outcome*

We found no evidence for a step change in renal recovery rate (return to a serum creatinine concentration within 120% of the baseline) following the intervention at the RFH. The estimated odds ratio (OR) for the intervention step change was 1.00 (95% CI 0.58-1.71). There was also no evidence for a significant difference in step change of recovery rate between RFH and BGH (estimated OR 1.24, 95% CI 0.53-2.92; *P*=.62).

The model did not estimate a statistically significant change in the trend of renal recovery rates at RFH (estimated OR 0.99, 95% CI 0.96-1.03; *P*=.61), indicating that the trend in the intervention period at RFH was not significantly different to that in the preintervention period. There was also no significant difference in the trend change between sites (estimated OR 0.97, 95% CI 0.92-1.03; *P*=.29). The data and model predictions are illustrated in Figure 2. Model estimates from the sensitivity analysis controlling for differences in casemix did not differ substantially from the primary analysis model estimates (Multimedia Appendix 1), and none of the 4 examined estimated ORs were statistically significantly different from 1.

**Figure 2.** Weekly recovery rate at Royal Free Hospital (RFH) and Barnet General Hospital (BGH) before and after implementation of the care pathway. Individual data points reflect the rate of each outcome for a single week. Solid lines indicate fitted values from the modeling functions.



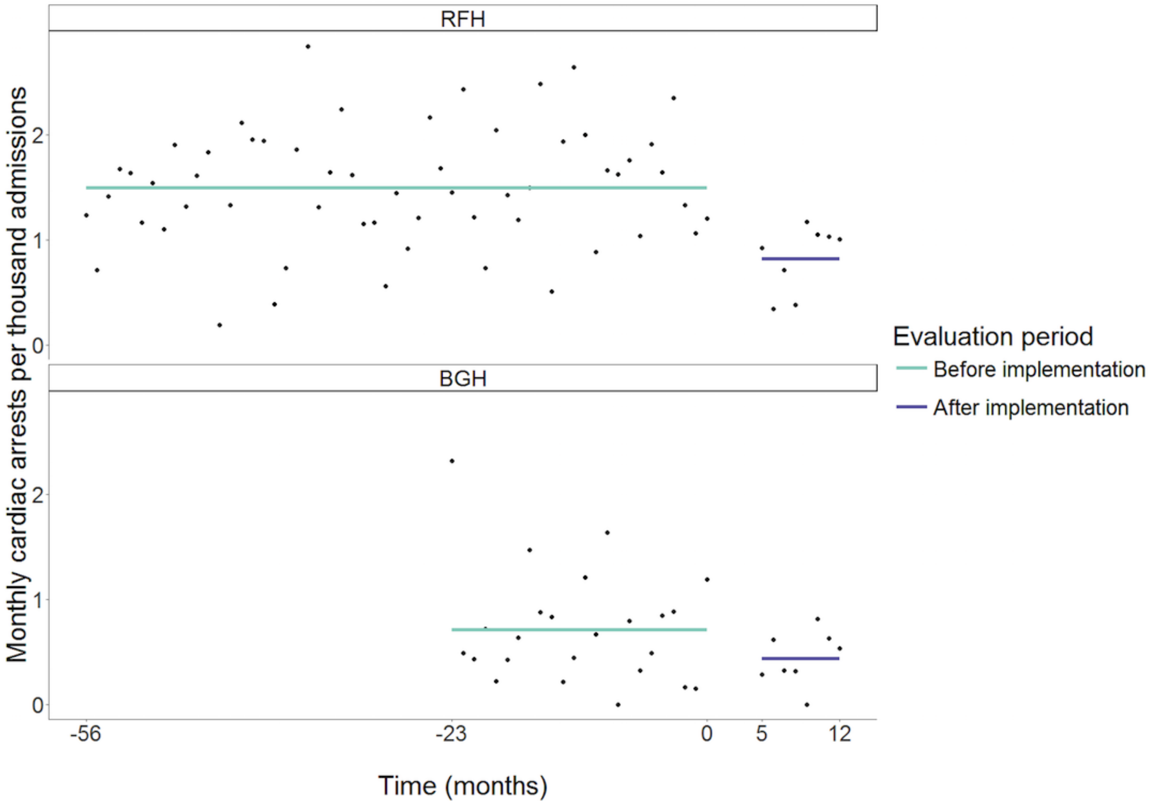

### *Secondary Clinical Outcomes*

We found evidence for a reduction (step change) in the rate of cardiac arrest following the intervention at RFH (estimated OR 0.55, 95% CI 0.38-0.76; *P*<.001). However, we found no statistically significant difference in the step change between sites (OR 1.13, 95% CI 0.63-1.99; *P*=.69). The data and model predictions are shown in Figure 3.

We also found evidence for a reduction (step change) in the rates of RRT use at 30 days at the intervention site (estimated OR 0.04, 95% CI 0.00-0.62, *P*=.04). However, because RRT was a rare event, estimates for this outcome were not reliable (Tables 4-7 and Multimedia Appendix 1). For all other secondary outcomes, models did not provide statistically significant evidence for an impact of the intervention. The data and model predictions are shown in Multimedia Appendix 1.

We found no evidence for an effect of the intervention on time to renal recovery. At RFH, the median (IQR) time to renal recovery was 3.00 days (1.00-15.00 days) before and 4.00 days (1.00-12.00 days) after the introduction of the intervention (*P*=.61). At BGH, the median (IQR) time to renal recovery was 3.00 (1.00-13.00) and 3.00 (1.00-7.00) days, respectively (*P*=.100). Using competing risk analysis, a significant increase in length of stay was demonstrated at both RFH (*P*=.046) and BGH (*P*=.03) after implementation of the care pathway (Multimedia Appendix 1).

**Figure 3.** Cardiac arrests at Royal Free Hospital (RFH) and Barnet General Hospital (BGH). Individual data points reflect the rate of cardiac arrest per thousand admissions for a single month. Solid lines indicate fitted values from the modeling functions.




## Economic Outcomes

There was a significant reduction in adjusted mean costs per spell over time at the RFH but not at the BGH (Tables 8-10). There was a significant reduction in mean costs per spell at the RFH in the postimplementation period compared with the preintervention period over and above the (nonsignificant) change seen at the BGH: the DID was −£2123 per spell (95% CI=−£4024 to −£222, *P*=.03). For the specified secondary analysis including all periods, the DID was −£1631 per spell (95% CI=−£3218 to −£44, *P*=.04). No significant differences were noted in the analyses of the cost components (Multimedia Appendix 1).

**Table 8.** Results of economic analysis: Royal Free Hospital.

| Time period | Preintervention (£) | | Postintervention (£) | | Difference (£) | | *P* value |
|---|---|---|---|---|---|---|---|
| | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | |
| Periods t1[a] and t3[b] only | 12,176.52 | 10,996.53 to 13,356.50 | 9853.37 | 8840.91 to 10,865.82 | −2323.15 | −3843.90 to −802.41 | .003 |
| All periods | 11,772.63 | 10,936.03 to 12,609.23 | 9761.59 | 8755.45 to 10,767.72 | −2011.05 | −3283.53 to −738.56 | .002 |

[a]t1: May to September 2016.

[b]t3: May to September 2017.

**Table 9.** Results of economic analysis: Barnet General Hospital.

| Time period | Preintervention (£) | | Postintervention (£) | | Difference (£) | | *P* value |
|---|---|---|---|---|---|---|---|
| | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | |
| Periods t1[a] and t3[b] only | 7507.88 | 6589.77 to 8425.99 | 7307.27 | 6461.82 to 8152.71 | −200.62 | −1370.27 to 969.04 | .74 |
| All periods | 7623.76 | 7007.67 to 8239.86 | 7243.58 | 6413.81 to 8073.35 | −380.19 | −1358.56 to 598.19 | .45 |

[a]t1: May to September 2016.

[b]t3: May to September 2017.

**Table 10.** Results of economic analysis: difference-in-difference analysis of Royal Free Hospital and Barnet General Hospital.

| Time period | Mean | 95% CI | *P* value |
|---|---|---|---|
| Periods t1[a] and t3[b] only (£) | −2122.54 | −4023.37 to −221.70 | .03 |
| All periods (£) | −1630.86 | −3217.50 to −44.22 | .04 |

[a]t1: May to September 2016.

[b]t3: May to September 2017.

# Discussion

## Principal Findings

The digitally enabled care pathway for the management of AKI in a large, acute hospital with a complex casemix resulted in no significant impact on the primary outcome of renal recovery or any of the other secondary clinical outcomes measured but was associated with a significant reduction in adjusted mean costs per patient admission. We did not include the costs of providing the technology, and therefore, it is not possible to judge whether or not it would be cost saving overall. Our results suggest that the digitally enabled care pathway would be cost saving, provided provision of the technology costs less than around £1600 per patient spell. The causes of the cost savings are unclear but are likely to be multifactorial, and further research to investigate these would be useful. The most important cost components contributing to this reduction (detailed in Multimedia Appendix 1) were length of stay and theater cutting time (which might itself be expected to play a role among patients requiring surgical intervention for AKI). There was a statistically significant reduction in the need for RRT at 30 days post-AKI; however, our model was not sufficiently reliable, given the low observed event rate of this outcome. The reduction in cardiac arrest rate needs to be viewed with caution because of the large number of hypothesis tests we conducted for our 6 prespecified secondary outcomes, and because this was a hospital-wide measure, this may have been influenced by other concurrently implemented initiatives. Furthermore, cardiac arrest rates also reduced at the comparator site. It is possible that both the RFH digital pathway and BGH's quality improvement initiative were effective to some extent through different mechanisms.

There are several possible explanations for the lack of impact on renal recovery. First, this may reflect existing high standards of AKI care before implementation: 30-day mortality for preintervention patients at RFH was 14.9% compared with 18.1% nationally [27]. It is possible that our intervention may have delivered more benefit in hospitals with worse outcomes. Second, AKI arising during inpatient admission has been shown to have worse outcomes than that arising at emergency presentation [28]. This may be because AKI arising during hospital treatment may be harder to modify. Third, AKI detection using NHSEDA depends on an elevation of serum creatinine, the detection of which may lag many hours or even days after the time of renal insult [29]. In consequence, renal injury may be less modifiable by this stage, even using a rapid system of detection such as that described. Finally, it is possible that the Streams app may have had a greater impact were it to have been implemented as part of a different care pathway—perhaps, one that involved general physicians as well as specialty care.

An explanation for the possible effect of the intervention on rates of cardiac arrest emerged from qualitative data provided in our parallel paper [30]. Here, users suggested the care pathway not only enhanced early access to specialist care for deteriorating patients but also informed treatment escalation plans. The latter included institution of ceilings of care and *do not resuscitate* orders with patients and relatives. Both would be expected to contribute to a reduction in the recorded unexpected cardiac arrest rate.

## Comparison With Prior Work

Our data are consistent with recent reports of the benefits of e-alerting systems for AKI for patients and the wider health system. We have reported on the impact of the digitally enabled care pathway on processes of care and clinical outcomes for patients with AKI at the point of presentation to the ED. Implementation of the digitally enabled care pathway for these patients was associated with significant improvement in the reliability of AKI recognition, a reduction in time to recognize and adjust potentially nephrotoxic medications [18]. Our qualitative analysis [30] found that care pathway improved access to patient information and expedited early specialist care. Our results concur with other research findings: a recent study from Korea [31] suggested that e-alerting for inpatients improves AKI recognition and the number of patients receiving specialist review [31]. Moreover, 2 single-site quality improvement projects combining AKI alerts with care bundles and targeted staff education also improved recognition of AKI and the quality of inpatient care [32,33]. In addition, a large multicenter sequential period analysis of an alerting system warning clinicians of the possible presence of AKI next to the display of serum creatinine results resulted in a small but sustained decrease in in-hospital mortality, dialysis use, and length of stay [34]. However, similar to our research, it is unclear which components of these pathways influenced these outcomes. A number of mixed-methods analyses of e-alerting systems for AKI are still underway; results from the qualitative segments of the AKORDD [35] and TACKLING [36] studies are awaited.

## Strengths and Limitations

Our evaluation had a number of strengths. First, this is, to our knowledge, the first study to define the economic impact of implementing a digital innovation for AKI on health systems. Second, we clinically validated all NHSEDA AKI alerts before analysis and validated this process. Third, our inclusion of a comparator site follows best practice [37], ensuring transparency in the drawing of conclusions about the *active* components of our intervention.

Our evaluation also had several limitations. First, longer time frames and the inclusion of multiple intervention and comparator sites would have helped overcome the effect of differences in casemix in the pre- and postintervention period (identified in our single comparator site) and may have helped to clarify any added value of the integration of our digital innovation into the care pathway. This would also have allowed us to investigate the impact on specific patient subgroups and better understand if outcomes differed between different AKI stages. It is possible, for instance, that established severe AKI is far less responsive to intervention than the early disease. It is important that such issues are prospectively addressed in future studies. Longer time frames would also have allowed us to control for any seasonal changes in outcome, which are known to occur [38] and should be borne in mind in the design of future studies. It was not possible to collect cost data relating to the innovation of the intervention site, which should be included in any future cost-benefit analyses. Finally, although time to in-app AKI recognition and virtual review by a specialist was very rapid (median 14.0 min), comparable data from the preimplementation phase could not be collected as this process is integral to the Streams app.

### Conclusions

The digitally enabled AKI care pathway reduced inpatient health care costs and may also help reduce hospital-wide cardiac arrest rates: this result requires reanalysis in larger, multisite studies. Growing support for greater digitalization of health systems offers the opportunity to improve the quality and safety of care and to reduce its cost. However, prospective evaluation of the clinical and cost impacts of digital innovations within the context in which they are delivered will be key in delivering maximum utility for patients and health systems.

## Acknowledgments

ECB and RR are supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care North Thames at Barts Health NHS Trust. RR is an NIHR Senior Investigator. GR and HM are funded in part by the NIHR University College London Hospitals Biomedical Research Centre. OSA is partially supported by an NIHR academic clinical fellowship. The authors wish to express their gratitude to the staff and patients of RFLFT. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

## Authors' Contributions

HM, CL, RR, CH, AK, TB, KA, DK, and MS initiated the project and collaboration. CL conceived the care pathway. AC, CL, CM, JC, GJ, SS, and ME supported implementation. AC, CL, GR, RR, HM, PM, and CN were responsible for the design of the evaluation. AC and CL triaged alerts necessary for the evaluation. AC collected all necessary data. Clinical outcomes were analyzed by AC with assistance and oversight from PM and CN. Economic outcomes were analyzed by ECB with assistance from SM. AC, HM, RR, PM, CL, ECB, SM, and GR wrote the paper. All authors read and agreed the final submission.

## Conflicts of Interest

CL, HM, GR, and RR are paid clinical advisors to DeepMind. AC's clinical research fellowship was part-funded by DeepMind, where he has been a full-time employee since May 2018. DeepMind remained independent from the collection and analysis of all data. CL was a member of the NICE clinical guideline 169 development group referenced in the article. HM coholds a patent on a fluid delivery device, which might ultimately help in preventing some (dehydration-related) cases of AKI occurring.

DeepMind was acquired by Google in 2014 and is now part of the Alphabet group. The deployment of Streams app at RFH was the subject of an investigation by the Information Commissioner's Office in 2017. RFH has since published an audit completed to comply with undertakings following this investigation [39]. In November 2018, it was announced that the Streams app team will be joining Google as part of a wider health effort [40].

## Multimedia Appendix 1

The National Health Service Early Detection Algorithm; acute kidney injury (AKI) care pathway; AKI care protocol; nursing advisory sticker; inter- and intra-operator variability analyses; Royal Free Hospital Data Monitoring Committee; distribution of costs per spell; results of segmented regression analyses; results from binary logistic regression analysis; graphs of secondary outcome data; cost component analyses.

[PDF File (Adobe PDF File), 1MB - jmir_v21i7e13147_app1.pdf ]

## References

1. Bellomo R, Kellum JA, Ronco C. Acute kidney injury. Lancet 2012 Aug 25;380(9843):756-766. [doi: 10.1016/S0140-6736(11)61454-2] [Medline: 22617274]
2. Zeng X, McMahon GM, Brunelli SM, Bates DW, Waikar SS. Incidence, outcomes, and comparisons across definitions of AKI in hospitalized individuals. Clin J Am Soc Nephrol 2014 Jan;9(1):12-20 [FREE Full text] [doi: 10.2215/CJN.02730313] [Medline: 24178971]

3. Metcalfe W, Simpson M, Khan IH, Prescott GJ, Simpson K, Smith WC, Scottish Renal Registry. Acute renal failure requiring renal replacement therapy: incidence and outcome. QJM 2002 Sep;95(9):579-583. [Medline: 12205335]
4. Wang HE, Muntner P, Chertow GM, Warnock DG. Acute kidney injury and mortality in hospitalized patients. Am J Nephrol 2012;35(4):349-355 [FREE Full text] [doi: 10.1159/000337487] [Medline: 22473149]
5. Chawla LS, Amdur RL, Amodeo S, Kimmel PL, Palant CE. The severity of acute kidney injury predicts progression to chronic kidney disease. Kidney Int 2011 Jun;79(12):1361-1369 [FREE Full text] [doi: 10.1038/ki.2011.42] [Medline: 21430640]
6. Connell A, Laing C. Acute kidney injury. Clin Med (Lond) 2015 Dec;15(6):581-584. [doi: 10.7861/clinmedicine.15-6-581] [Medline: 26621953]
7. Kerr M, Bedford M, Matthews B, O'Donoghue D. The economic impact of acute kidney injury in England. Nephrol Dial Transplant 2014 Jul;29(7):1362-1368. [doi: 10.1093/ndt/gfu016] [Medline: 24753459]
8. Silver S, Chertow G. The economic consequences of acute kidney injury. Nephron 2017;137(4):297-301 [FREE Full text] [doi: 10.1159/000475607] [Medline: 28595193]
9. National Institute for Health and Care Excellence. Acute kidney injury: prevention, detection and management URL: https://www.nice.org.uk/guidance/cg169 [accessed 2018-11-14] [WebCite Cache ID 74eveiIbK]
10. Alleway R. NCEPOD. 2009. Acute Kidney Injury: Adding Insult to Injury Report URL: http://www.ncepod.org.uk/2009aki.html [accessed 2016-03-08]
11. KDIGO working group. KDIGO. 2012. KDIGO Clinical practice guidelines for acute kidney injury URL: https://kdigo.org/wp-content/uploads/2016/10/KDIGO-2012-AKI-Guideline-English.pdf
12. Wilson FP, Shashaty M, Testani J, Aqeel I, Borovskiy Y, Ellenberg SS, et al. Automated, electronic alerts for acute kidney injury: a single-blind, parallel-group, randomised controlled trial. Lancet 2015 May 16;385(9981):1966-1974. [doi: 10.1016/S0140-6736(15)60266-5] [Medline: 25726515]
13. NHS England. 2016. Patient safety alert on standardising the early identification of Acute Kidney Injury URL: https://www.england.nhs.uk/2014/06/psa-aki/ [accessed 2016-03-08]
14. Lachance P, Villeneuve P, Rewa OG, Wilson FP, Selby NM, Featherstone RM, et al. Association between e-alert implementation for detection of acute kidney injury and outcomes: a systematic review. Nephrol Dial Transplant 2017 Dec 1;32(2):265-272 [FREE Full text] [doi: 10.1093/ndt/gfw424] [Medline: 28088774]
15. Master J, Hammad S, Chamberlain P, Chandrasekar T, Wong C. Reduction in acute kidney injury (AKI) mortality data with the development of a novel AKI management bundle. Nephrol Dial Transplant 2015;30:iii467. [doi: 10.1093/ndt/gfv190.80]
16. Chandrasekar T, Sharma A, Chamberlain P, Abraham A. streamlined approach to AKI management leads to reduction of in-hospital mortality, length of stay and AKI progression. Nephrol Dial Transplant 2016;31:170-171. [doi: 10.1093/ndt/gfw147.03]
17. Connell A, Montgomery H, Morris S, Nightingale C, Stanley S, Emerson M, et al. Service evaluation of the implementation of a digitally-enabled care pathway for the recognition and management of acute kidney injury. F1000Res 2017;6:1033. [doi: 10.12688/f1000research.11637.2] [Medline: 28751970]
18. Connell A, Montgomery H, Martin P, Nightingale C, Sadeghi-Alavijeh O, King D, et al. Evaluation of a digitally-enabled care pathway for the management of acute kidney injury in patients admitted to hospital as an emergency (in press). 2019.
19. Thygesen S, Christiansen C, Christensen S, Lash T, Sørensen H. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med Res Methodol 2011 May 28:11. [Medline: 21619668]
20. IMD by postcode lookup. URL: http://imd-by-postcode.opendatacommunities.org/ [accessed 2018-07-23]
21. NHS Data Dictionary. Attributes URL: https://www.datadictionary.nhs.uk/data_dictionary/attributes/attributes.asp [accessed 2018-08-14]
22. Sawhney S, Fluck N, Marks A, Prescott G, Simpson W, Tomlinson L, et al. Acute kidney injury: how does automated detection perform? Nephrol Dial Transplant 2015 Nov 1;30(11):1853-1861. [doi: 10.1093/ndt/gfv094]
23. R Foundation for Statistical Computing. 2017. URL: http://www.R-project.org/
24. Stata: Software for Statistics and Data Science. URL: https://www.stata.com/
25. Brock G, Barnes C, Ramirez J, Myers J. How to handle mortality when investigating length of hospital stay and time to clinical stability. BMC Med Res Methodol 2011 Oct 26;11:144. [Medline: 22029846]
26. NHS Health Research Authority. Is my study research? URL: http://www.hra-decisiontools.org.uk/research/redirect.html [accessed 2018-03-23]
27. Think Kidneys. Reporting the rate of Acute Kidney Injury (AKI) within England: the current state of the NHS AKI Master Patient Index and Registry URL: https://www.thinkkidneys.nhs.uk/aki/wp-content/uploads/sites/2/2018/03/Reporting-the-rate-of-AKI-January-2018.pdf [accessed 2018-03-30]
28. Hsu C, Lee C, Su C, Wang Y, Chen H, Chuang J, et al. Incidence, outcomes, and risk factors of community-acquired and hospital-acquired acute kidney injury: a retrospective cohort study. Medicine 2016 May;95(19):e3674. [Medline: 27175701]
29. Thomas M, Blaine C, Dawnay A, Devonald M, Ftouh S, Laing C, et al. The definition of acute kidney injury and its use in practice. Kidney Int 2015 Jan;87(1):62-73 [FREE Full text] [doi: 10.1038/ki.2014.328] [Medline: 25317932]

30. Connell A, Black G, Montgomery H, Martin P, Nightingale C, King D, et al. Implementation of a Digitally Enabled Care Pathway (Part 2): Qualitative Analysis of Experiences of Healthcare Professionals. J Med Internet Res 2019 Jul 31;21(7):e13143. [doi: 10.2196/13143]
31. Park S, Baek S, Ahn S, Lee K, Hwang H, Ryu J, et al. Impact of electronic acute kidney injury (AKI) alerts with automated nephrologist consultation on detection and severity of AKI: a quality improvement study. Am J Kidney Dis 2017 Jul 25;71(1):9-19. [Medline: 28754457]
32. Ebah L, Hanumapura P, Waring D, Challiner R, Hayden K, Alexander J, et al. A multifaceted quality improvement programme to improve acute kidney injury care and outcomes in a large teaching hospital. BMJ Qual Improv Rep 2017 May 25;6(1). [Medline: 28607684]
33. Chandrasekar T, Sharma A, Tennent L, Wong C, Chamberlain P, Abraham KA. A whole system approach to improving mortality associated with acute kidney injury. QJM 2017 Oct 1;110(10):657-666. [doi: 10.1093/qjmed/hcx101] [Medline: 28521019]
34. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical decision support for in-hospital AKI. J Am Soc Nephrol 2018 Dec;29(2):654-660 [FREE Full text] [doi: 10.1681/ASN.2017070765] [Medline: 29097621]
35. Abdelaziz TS, Lindenmeyer A, Baharani J, Mistry H, Sitch A, Temple RM, et al. Acute Kidney Outreach to Reduce Deterioration and Death (AKORDD) trial: the protocol for a large pilot study. BMJ Open 2016 Dec 19;6(8):e012253 [FREE Full text] [doi: 10.1136/bmjopen-2016-012253] [Medline: 27543592]
36. Selby NM, Casula A, Lamming L, Mohammed M, Caskey F, Tackling AKI Investigators. Design and rationale of 'tackling acute kidney injury', a multicentre quality improvement study. Nephron 2016;134(3):200-204 [FREE Full text] [doi: 10.1159/000447675] [Medline: 27376867]
37. Raine R, Fitzpatrick R, Barratt H, Bevan G, Black N, Boaden R, et al. National Institute for Health Research. Southampton (UK): NIHR Journals Library; 2016. Challenges, solutions and future directions in the evaluation of service innovations in health care and public health URL: https://njl-admin.nihr.ac.uk/document/download/2004136
38. Phillips D, Young O, Holmes J, Allen LA, Roberts G, Geen J, Welsh AKI steering group. Seasonal pattern of incidence and outcome of acute kidney injury: a national study of Welsh AKI electronic alerts. Int J Clin Pract 2017 Sep;71(9). [doi: 10.1111/ijcp.13000] [Medline: 28869717]
39. Royal Free London NHS Foundation Trust. Royal Free London publishes audit into Streams app URL: https://www.royalfree.nhs.uk/news-media/news/royal-free-london-publishes-audit-into-streams-app/ [accessed 2018-10-18]
40. DeepMind Health. Scaling Streams with Google URL: https://deepmind.com/blog/scaling-streams-google/ [accessed 2018-11-21]

## Abbreviations

**AKI:** acute kidney injury
**BGH:** Barnet General Hospital
**DID:** difference-in-differences
**ED:** emergency department
**GLM:** generalized linear model
**IMD:** Index of Multiple Deprivation
**IQR:** interquartile range
**ITU:** intensive treatment unit
**NHS:** National Health Service
**NHSEDA:** NHS Early Detection Algorithm
**NIHR:** National Institute for Health Research
**OR:** odds ratio
**PARRT:** Patient at Risk and Resuscitation Team
**PLICS:** Payment Level Information and Costing System
**RFH:** Royal Free Hospital
**RFLFT:** Royal Free London NHS Foundation Trust
**RRT:** renal replacement therapy
**UCL:** University College London

*Edited by G Eysenbach; submitted 14.12.18; peer-reviewed by D O'Donoghue, X Ji, T Kothari; comments to author 16.01.19; revised version received 29.01.19; accepted 30.01.19; published 15.07.19.*

*Please cite as:*
*Connell A, Raine R, Martin P, Barbosa EC, Morris S, Nightingale C, Sadeghi-Alavijeh O, King D, Karthikesalingam A, Hughes C, Back T, Ayoub K, Suleyman M, Jones G, Cross J, Stanley S, Emerson M, Merrick C, Rees G, Montgomery H, Laing C*
*Implementation of a Digitally Enabled Care Pathway (Part 1): Impact on Clinical Outcomes and Associated Health Care Costs*
*J Med Internet Res 2019;21(7):e13147*
*URL: http://www.jmir.org/2019/7/e13147/*
*doi:10.2196/13147*
*PMID:31368447*

Original Paper

# Online Ratings of Urologists: Comprehensive Analysis

C William Pike[1], BA; Jacqueline Zillioux[2], MD; David Rapp[2], MD

[1]Georgetown University School of Medicine, Washington, DC, United States

[2]Department of Urology, University of Virginia Medical Center, Charlottesville, VA, United States

**Corresponding Author:**
David Rapp, MD
Department of Urology
University of Virginia Medical Center
500 Ray C Hunt Drive
Charlottesville, VA, 22908
United States
Phone: 1 8043859511
Email: derapp@yahoo.com

## Abstract

**Background:** Physician-rating websites are being increasingly used by patients to help guide physician choice. As such, an understanding of these websites and factors that influence ratings is valuable to physicians.

**Objective:** We sought to perform a comprehensive analysis of online urology ratings information, with a specific focus on the relationship between number of ratings or comments and overall physician rating.

**Methods:** We analyzed urologist ratings on the Healthgrades website. The data retrieval focused on physician and staff ratings information. Our analysis included descriptive statistics of physician and staff ratings and correlation analysis between physician or staff performance and overall physician rating. Finally, we performed a best-fit analysis to assess for an association between number of physician ratings and overall rating.

**Results:** From a total of 9921 urology profiles analyzed, there were 99,959 ratings and 23,492 comments. Most ratings were either 5 ("excellent") (67.53%, 67,505/99,959) or 1 ("poor") (24.22%, 24,218/99,959). All physician and staff performance ratings demonstrated a positive and statistically significant correlation with overall physician rating ($P$<.001 for all analyses). Best-fit analysis demonstrated a negative relationship between number of ratings or comments and overall rating until physicians achieved 21 ratings or 6 comments. Thereafter, a positive relationship was seen.

**Conclusions:** In our study, a dichotomous rating distribution was seen with more than 90% of ratings being either excellent or poor. A negative relationship between number of ratings or comments and overall rating was initially seen, after which a positive relationship was demonstrated. Combined, these data suggest that physicians can benefit from understanding online ratings and that proactive steps to encourage patient rating submissions may help optimize overall rating.

***(J Med Internet Res 2019;21(7):e12436)*** doi:10.2196/12436

**KEYWORDS**

online physician ratings; urology; reputation management

## Introduction

Recent data demonstrate that most Americans use the internet to search for health information [1-3]. In addition, a large percentage of patients obtain information about physicians through internet resources and identify online websites as important in their choice of health care providers [4,5]. A prior study evaluating patient trends reported that 59% of the US population reported physician-rating websites (PRWs) to be somewhat important in choosing their health care providers [4]. At the same time, there has been a tremendous growth in the number of PRWs [6]. There exist at least 28 PRWs that display information about physician training and allow users to rate physician or staff characteristics.

Although criticisms regarding the validity of PRWs are often raised by physicians, these data show the importance of online physician reputations. In addition to the importance of PRWs in guiding patient selection as consumers, online rating systems are also part of a more widespread focus on the patient experience. Accordingly, the Hospital Quality Alliance was established in an effort to promote transparency of care quality reporting [7]. The initiatives of the Hospital Quality Alliance and Medicare are seen in publicly available data focused on core care measures that include patient surveys about their care (Hospital Consumer Assessment of Healthcare Providers and Systems). As such measures of patient experience become more commonly used to assess care quality and influence reimbursement models (eg, value-based purchasing), it becomes even more important that physicians maintain a working knowledge of online patient reviews.

Even so, investigation suggests that many physicians have little familiarity with PRW, do not commonly check their own reviews, and spend minimal time managing their digital reputation [8]. Although reputation management is a frequent focus in commerce and marketing literature, little is written about online reputation management of physicians. Suggestions for optimization of online ratings within the general literature include actively encouraging patients to submit ratings and responding to negative comments online [9].

Within the urology literature, we could identify only two studies focused on the assessment of online ratings [10,11]. Thus, we sought to comprehensively assess online ratings in a large cohort of urologists. Specific study aims included the assessment of the relationship between number of ratings and the overall mean rating. We hypothesized that number of ratings would demonstrate a positive correlation with overall ratings as this may reflect initiatives by certain physicians to actively manage their reputation and encourage patients to submit online comments or ratings. We also sought to assess the distribution of ratings and assess for a correlation between individual physician and staff characteristic ratings and overall rating.

## Methods

We conducted an analysis of urologic physician ratings and related information on the website Healthgrades. Data retrieval was facilitated using Java (version 8). Specific focus was placed on aggregating data related to physician and staff ratings, including number and distribution of ratings, number of comments, physician performance characteristics, as well as office and staff performance characteristics.

In brief, overall physician ratings are provided as a score between 1 and 5 (1=poor; 5=excellent). Ratings are also available for a specific physician (trustworthiness, explains conditions well, answers questions, time well spent) and staff (scheduling, office environment, staff friendliness) performance variables. Each of these physician and staff performance variables is also rated on a score of 1 to 5.

Inclusion and exclusion criteria were designed in an effort to focus on a cohort of actively practicing urologists and exclude those that may be in residency, retired, or deceased. Accordingly, physicians with a known age of 35 to 74 years were included. These age criteria were selected after an initial data review of age-related ratings distribution, which revealed that the majority of physicians with ages younger than 35 or older than 74 years had zero ratings. Physicians without data specifically detailing age or an age estimation (years out from medical school) were also excluded.

Analysis first focused on descriptive statistics to assess overall physician rating, number of ratings or comments per physician, and ratings related to specific physician and staff performance variables. Variables are presented as mean and standard deviation. We then assessed for a Pearson correlation between physician and staff performance variables and overall physician rating. Finally, we performed a best-fit analysis to assess for an association between number of physician ratings or comments and overall rating. Statistical analysis was performed using R (version 3.4.1). All tests were performed with $\alpha$=0.05. The University of Virginia (Charlottesville, VA) institutional review board determined that this study met the criteria for nonhuman research (IRB #: 20592).

## Results

Data were retrieved for 14,430 urologists, of which 9921 met the inclusion criteria and were included in study analysis. A total of 99,959 ratings and 23,492 comments were seen across 9921 urologists. The mean number of ratings and comments per urologist was 10.1 (SD 4.3) and 2.4 (SD 6.0), respectively. In addition, a significant range in number of ratings (0-395) and comments (0-241) per urologist was seen. Analysis demonstrated that 1554 of 9921 (15.66%) and 4077 of 9921 (41.09%) of physicians had zero ratings and zero comments, respectively.

The distribution of ratings is showed in Figure 1. The vast majority of ratings were either 5 ("excellent") (67.53%, 67,505/99,959) or 1 ("poor") (24.22%, 24,218/99,959). Mean overall physician rating was 3.9 (SD 1.7). Physician and staff performance variable statistics and their correlation with overall ratings are detailed in Table 1. All physician and staff performance ratings demonstrated a positive and statistically significant correlation with overall physician rating. The statistical coefficients (*R* value) for trustworthiness and answers questions were highest. Physician measures had higher correlations with overall rating than did staff measures.

Best-fit analyses of the relationship between number of ratings and overall physician rating as well as between number of comments and overall physician rating are shown in Figures 2 and 3 with locally weighted smoothing added for clarity. A negative relationship between the number of ratings and overall rating was seen until physicians achieved 21 ratings; thereafter, a positive relationship was seen. Similarly, a U-shaped relationship was seen when assessing the relationship between number of comments and overall rating, with the transition point being six comments.

**Figure 1.** Distribution of ratings of urologists on a physician-rating website (N=99,959).

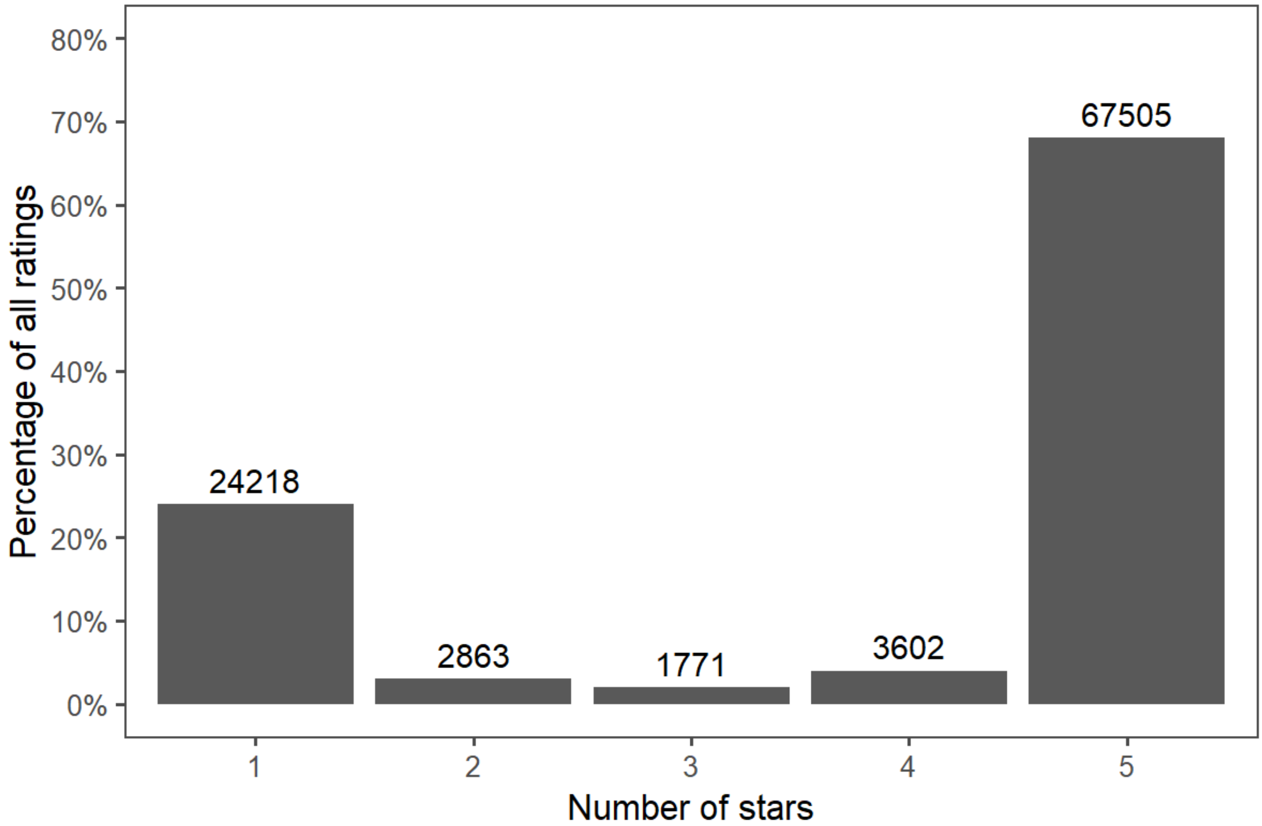



**Table 1.** Physician and staff performance variables and correlation with overall ratings.

| Performance variable | Rating, mean (SD) | Correlation with overall rating[a], *R* | *P* value |
|---|---|---|---|
| **Physician measures** | | | |
| Trustworthiness | 3.96 (0.89) | .965 | <.001 |
| Explains conditions well | 3.99 (0.88) | .953 | <.001 |
| Answers questions | 4.20 (0.70) | .957 | <.001 |
| Time well spent | 4.02 (0.76) | .947 | <.001 |
| **Staff measures** | | | |
| Scheduling | 4.09 (0.75) | .805 | <.001 |
| Office environment | 3.96 (0.88) | .796 | <.001 |
| Staff friendliness | 3.98 (0.88) | .817 | <.001 |

[a]Overall rating, mean (SD)=3.87 (1.72).

**Figure 2.** Relationship between number of ratings and overall physician rating.

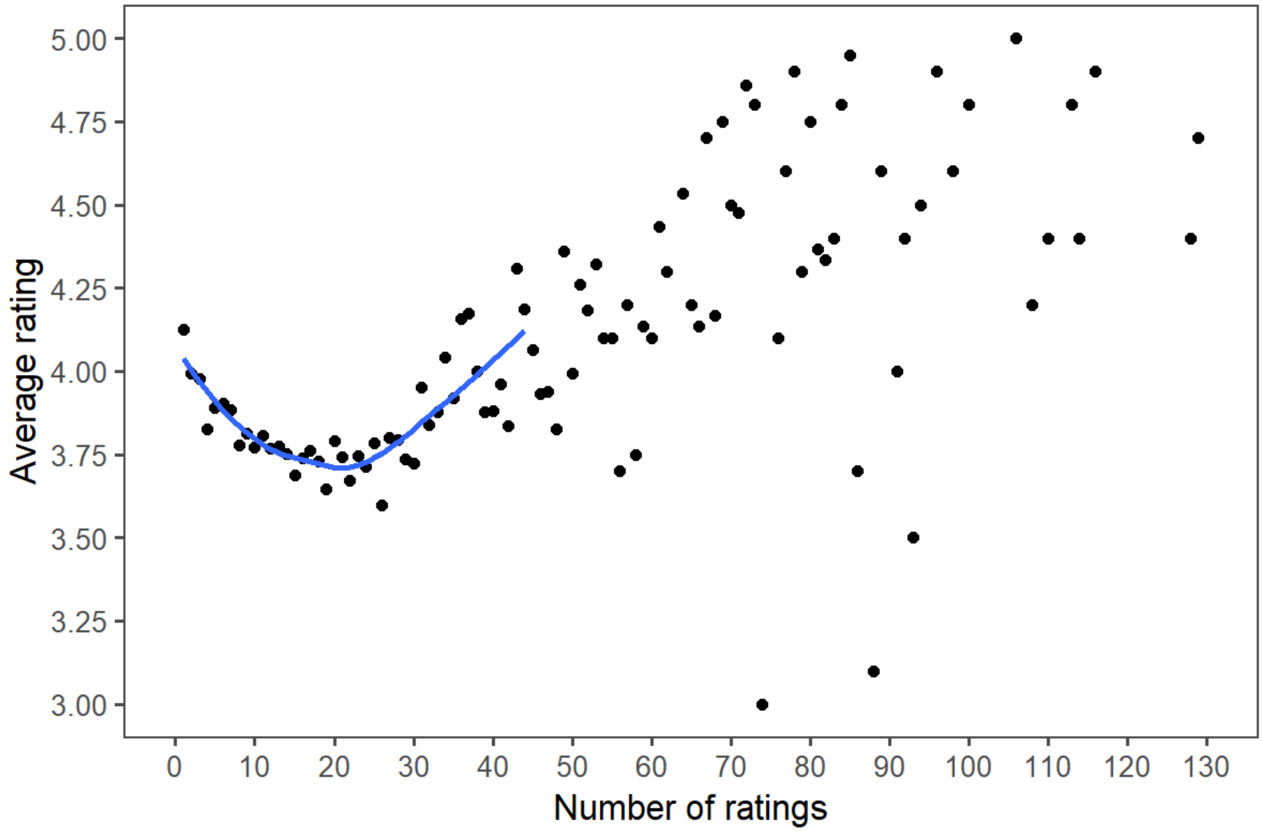
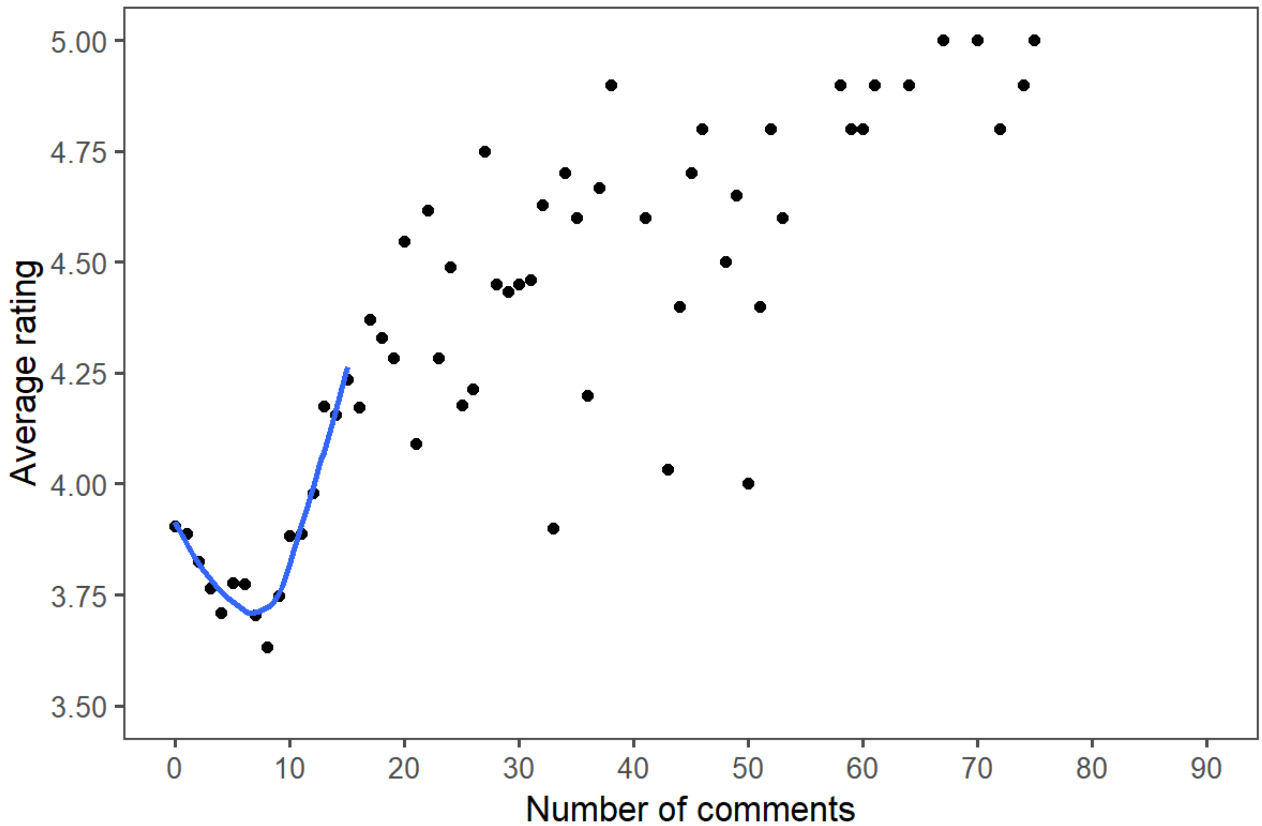



**Figure 3.** Relationship between number of comments and overall physician rating.



## Discussion

This study reveals several important findings. To our knowledge, this is the largest study comprehensively analyzing online ratings of a nationwide sample of urologists. Overall, more than 80% of urologists had at least one rating, demonstrating the use of PRWs by patients. These data are consistent with a prior study evaluating urologist ratings, which highlighted only that a small percentage of physicians do not have ratings associated with their profiles [10,11]. Interestingly, other investigations demonstrated that a large percentage of physicians overall have no ratings, but that specialists are twice as likely to have online ratings when compared with generalists [6,12]. Combined with our investigation, these data highlight the high utilization of PRWs within the surgical community and the need to demonstrate awareness with one's online ratings.

Second, the vast majority of ratings submitted were either excellent (5) or poor (1), with almost one-quarter of ratings being 1. Other studies have described ratings distributions, with most being positive. Kadry and colleagues [13] demonstrated that approximately two of three patient reviews were favorable across 23 specialties. Lagu et al [14] found that 88% of ratings were positive for both generalists and specialists (as defined by a rating of 3 or greater on a 4-point scale). A prior study also demonstrated that most comments on PRWs are positive [15]. Generally, our findings are consistent with these prior investigations suggesting that most ratings are either excellent or poor [16]. Notably, in review of physician ratings on RateMD.com, Gao et al [17] found that 42% of the ratings were between 2 and 4 on a 5-point scale. This differs significantly from our results, which demonstrate an extreme dichotomy of ratings.

Further, all performance variables assessed strongly correlated with overall physician rating. This finding suggests that these variables all influence a patient's overall satisfaction with the visit. A prior study demonstrated a statistically significant correlation between staff and physician ratings [17]. In addition, Kadry and colleagues [13] reported a strong correlation between a diverse number of dimensions of the patient appointment and overall rating. In this study, dimensions assessed included communication skills (eg, listens and answers questions) and access (eg, ease of appointment, punctuality). Our analysis adds to this literature because it assesses further variables that may influence patient satisfaction.

Most notably, our analysis demonstrated a U-shaped relationship between number of ratings and overall mean rating. A similar relationship was observed in the relationship between number of comments and overall mean rating. Accordingly, before achieving 21 patient ratings, a negative relationship was demonstrated between number of ratings and overall mean rating, followed by a positive relationship. Similarly, a rating nadir was seen at six comments, after which a positive relationship was seen. We hypothesize that this relationship is created from the significant impact that a single poor rating can have on the overall mean rating when there are few ratings. Prior opinion supports this theory, suggesting that in cases in which there are few ratings, one outlying value or comment can have a disproportionately large influence [18].

Combined, these findings emphasize the importance of active knowledge and management of online reputation by urologists. Experience related to online reputation management suggests that a single negative review likely has a greater influence than multiple positive evaluations [19]. Despite this fact, a large percentage of physicians do not check their online profile [8,20]. Physician criticism of PRWs is understandable given previous studies demonstrating inconsistencies between patient ratings and quality of care [21-25]. However, given data demonstrating the rapid increase in the utilization of PRWs by patients, it is important that physicians have a working knowledge of their online reputation. Further, patients are increasingly providing online reviews of hospitals and treatment centers [26]. As patient satisfaction with physicians can also be influenced by the patient experience at the hospitals where they offer care, physician awareness of these facility reviews is also important.

In addition, active steps by physicians should be considered to help optimize ratings. Foremost, our data suggest that efforts should focus on building total volume of reviews on PRWs. Suggested approaches involve the use of collateral to solicit reviews, including patient cards, videos, and emails [9]. In addition, patients can be encouraged to complete online ratings and surveys at the time of encounter, thus offering a more proactive approach [8]. Finally, appropriately addressing negative comments or providing personalized review responses is suggested as a potential method of demonstrating physician focus on the patient experience to other potential patients visiting the PRW [27]. This is important given a study showing that only 39% of physicians agree with their profile ratings [20]. Further study is ongoing at our institution to assess specific methods of optimizing patient engagement and ratings.

Beyond commercial PRWs, focus on additional online forums can help optimize physicians' digital reputations. One such method includes using online professional networking websites (eg, Doximity) to publish professional accomplishments [28]. The creation of a personal online blog by physicians offers another technique to share information with patients [8]. Further, the use of social media pages (eg, Facebook) can be an effective method of managing online reputation. Indeed, social media presence, such as Facebook followers, has been shown to correlate with US News and World Report reputation score [29]. Finally, utilizing noncommercial PRWs can also be valuable because the percentage of positive comments has been shown to be higher on health systems' online review websites when compared to commercial PRWs [15].

Study limitations include the study focus on ratings from a single PRW. Accordingly, the findings in this study may not be representative of trends across all PRWs. Healthgrades was selected because it is the most widely used PRW [13]. Supporting this trend is a prior systematic review showing that Healthgrades is the most widely selected PRW assessed in published investigations [26]. In addition, our study aim was to systematically assess ratings information across a large cohort of urologists through use of Java programming. Indeed, our cohort consisted of almost 10,000 urology profiles. However, given this methodology, analysis of text comments was not possible. Similarly, given the variability in rating scales and domains across the PRW, the inclusion of multiple PRWs and systematic comparison is difficult. Novel methods of automated analysis of text reviews have been recently reported and may allow for a more comprehensive study of text-based patient reviews in the future [30,31].

Nonetheless, we believe our study conclusions are strengthened by the large size of our cohort. Prior systematic review of studies on patient online reviews demonstrated that, in general, the number of providers with online reviews reported in investigations represented only a small percentage of the total workforce [26]. Recent data from the American Board of Medical Specialties reports 13,039 board-certified urologists [32]. Although there may be additional urologists without board certification, these data suggest that our study captured over 75% of all urologists within the United States and highlight the significance of our cohort size. In addition, our data represent a large and diverse sample size across various regions, practice types, physician ages, and other physician characteristics. Future study is ongoing to assess the potential relationship between these variables and online ratings. Such study is important given conflicting data regarding the relationship between physician characteristics (such as gender, practice experience, and academic productivity) and patient online ratings [26].

In conclusion, our study demonstrates that most online urologist profiles have received ratings. Further, a dichotomous rating distribution is seen, with more than 90% of ratings being either poor or excellent. A negative relationship between number of ratings and overall rating is initially seen, following which a positive relationship is demonstrated. Combined, these data suggest that physicians can benefit from understanding ratings associated with their online profile and that proactive steps to optimize their rating may be helpful.

## Conflicts of Interest

None declared.

## References

1. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]
2. Segal J. The role of the Internet in doctor performance rating. Pain Physician 2009;12(3):659-664 [FREE Full text] [Medline: 19461833]
3. The Harris Poll. 2010 Aug 04. "Cyberchondriacs" on the rise? URL: https://tinyurl.com/y3xkhqz5 [accessed 2019-05-29] [WebCite Cache ID 78jropFIF]
4. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
5. Kuehn BM. More than one-third of US individuals use the Internet to self-diagnose. JAMA 2013 Feb 27;309(8):756-757. [doi: 10.1001/jama.2013.629] [Medline: 23443421]
6. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. JAMA 2017 Dec 21;317(7):766-768 [FREE Full text] [doi: 10.1001/jama.2016.18553] [Medline: 28241346]
7. CMS.gov: Centers for Medicare & Medicaid Services. Hospital Compare URL: https://tinyurl.com/y5bfcqz2 [accessed 2019-05-31] [WebCite Cache ID 78mlP0HY1]
8. Waxer JF, Srivastav S, DiBiase CS, DiBiase SJ. Investigation of radiation oncologists' awareness of online reputation management. JMIR Cancer 2019 Apr 01;5(1):e10530 [FREE Full text] [doi: 10.2196/10530] [Medline: 30932863]
9. Baker T. Physicians Practice. 2015 Nov 13. Five online reputation management strategies for physicians URL: http://www.physicianspractice.com/marketing/five-online-reputation-management-strategies-physicians [accessed 2019-05-31] [WebCite Cache ID 78ml64KdY]
10. Asanad K, Parameshwar PS, Houman J, Spiegel BM, Daskivich TJ, Anger JT. Online physician reviews in female pelvic medicine and reconstructive surgery: what do patients really want? Female Pelvic Med Reconstr Surg 2018;24(2):109-114. [doi: 10.1097/SPV.0000000000000503] [Medline: 29474282]
11. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
12. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
13. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
14. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]

15. Lagu T, Norton CM, Russo LM, Priya A, Goff SL, Lindenauer PK. Reporting of patient experience data on health systems' websites and commercial physician-rating websites: mixed-methods analysis. J Med Internet Res 2019 Mar 27;21(3):e12007 [FREE Full text] [doi: 10.2196/12007] [Medline: 30916654]
16. Hu N, Pavlou P, Zhang J. Overcoming the J-shaped distribution of product reviews. Commun ACM 2009 Oct;52(10):e1 [FREE Full text]
17. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
18. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? JAMA Otolaryngol Head Neck Surg 2014 Jul;140(7):635-638. [doi: 10.1001/jamaoto.2014.818] [Medline: 24876073]
19. ReputationManagement.com. URL: https://www.reputationmanagement.com/ [accessed 2018-08-14]
20. Johnson C. Survey finds physicians very wary of doctor ratings. Physician Exec 2013;39(1):6-8, 10, 12. [Medline: 23437751]
21. Kennedy GD, Tevis SE, Kent KC. Is there a relationship between patient satisfaction and favorable outcomes? Ann Surg 2014 Oct;260(4):592-598; discussion 598 [FREE Full text] [doi: 10.1097/SLA.0000000000000932] [Medline: 25203875]
22. Tevis SE, Kennedy GD, Kent KC. Is there a relationship between patient satisfaction and favorable surgical outcomes? Adv Surg 2015;49:221-233 [FREE Full text] [doi: 10.1016/j.yasu.2015.03.006] [Medline: 26299501]
23. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
24. Campbell L, Li Y. Are Facebook user ratings associated with hospital cost, quality and patient satisfaction? A cross-sectional analysis of hospitals in New York State. BMJ Qual Saf 2018 Dec;27(2):119-129. [doi: 10.1136/bmjqs-2016-006291] [Medline: 28600452]
25. Trehan SK, Nguyen JT, Marx R, Cross MB, Pan TJ, Daluiski A, et al. Online patient ratings are not correlated with total knee replacement surgeon-specific outcomes. HSS J 2018 Jul;14(2):177-180. [doi: 10.1007/s11420-017-9600-6] [Medline: 29983660]
26. Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL. What do patients say about doctors Online? A systematic review of studies on patient online reviews. J Med Internet Res 2019 Apr 08;21(4):e12521 [FREE Full text] [doi: 10.2196/12521] [Medline: 30958276]
27. Emmert M, Sauter L, Jablonski L, Sander U, Taheri-Zadeh F. Do physicians respond to web-based patient ratings? An analysis of physicians' responses to more than one million web-based ratings over a six-year period. J Med Internet Res 2017 Jul 26;19(7):e275 [FREE Full text] [doi: 10.2196/jmir.7538] [Medline: 28747292]
28. Prabhu AV, Kim C, De Guzman E, Zhao E, Madill E, Cohen J, et al. Reputation management and content control: an analysis of radiation oncologists' digital identities. Int J Radiat Oncol Biol Phys 2017 Dec 01;99(5):1083-1091. [doi: 10.1016/j.ijrobp.2017.08.015] [Medline: 28939228]
29. Triemstra JD, Poeppelman RS, Arora VM. Correlations between hospitals' social media presence and reputation score and ranking: cross-sectional analysis. J Med Internet Res 2018 Nov 08;20(11):e289 [FREE Full text] [doi: 10.2196/jmir.9713] [Medline: 30409768]
30. Rivas R, Montazeri N, Le NX, Hristidis V. Automatic classification of online doctor reviews: evaluation of text classifier algorithms. J Med Internet Res 2018 Nov 12;20(11):e11141 [FREE Full text] [doi: 10.2196/11141] [Medline: 30425030]
31. Li J, Liu M, Li X, Liu X, Liu J. Developing embedded taxonomy and mining patients' interests from web-based physician reviews: mixed-methods approach. J Med Internet Res 2018 Dec 16;20(8):e254 [FREE Full text] [doi: 10.2196/jmir.8868] [Medline: 30115610]
32. ABMS Board Certification Report: 2016-2017. Chicago, IL: American Board of Medical Specialties; 2017. URL: http://www.abms.org/media/139572/abms_board_certification_report_2016-17.pdf [accessed 2019-05-31] [WebCite Cache ID 78ml90KuM]


## Abbreviations

**PRW:** physician-rating website


*Edited by G Eysenbach; submitted 06.10.18; peer-reviewed by K Zheng, R Smith; comments to author 05.03.19; revised version received 23.03.19; accepted 02.05.19; published 02.07.19.*

*Please cite as:*
*Pike CW, Zillioux J, Rapp D*
*Online Ratings of Urologists: Comprehensive Analysis*
*J Med Internet Res 2019;21(7):e12436*
*URL: https://www.jmir.org/2019/7/e12436/*
*doi:10.2196/12436*
*PMID:31267982*

Original Paper

# Quantitative Ratings and Narrative Comments on Swiss Physician Rating Websites: Frequency Analysis

Stuart McLennan[1,2], MBHL, PhD

[1]Institute of History and Ethics in Medicine, Technical University of Munich, Munich, Germany

[2]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

**Corresponding Author:**
Stuart McLennan, MBHL, PhD
Institute of History and Ethics in Medicine
Technical University of Munich
Ismaninger Straße 22
Munich, 81675
Germany
Phone: 49 089 4140 4041
Fax: 49 089 4140 4970
Email: stuart.mclennan@tum.de

## Abstract

**Background:** Physician rating websites (PRWs) have been developed as part of a wider move toward transparency around health care quality, and these allow patients to anonymously rate, comment, and discuss physicians' quality on the Web. The first Swiss PRWs were established in 2008, at the same time as many international PRWs. However, there has been limited research conducted on PRWs in Switzerland to date. International research has indicated that a key shortcoming of PRWs is that they have an insufficient number of ratings.

**Objective:** The aim of this study was to examine the frequency of quantitative ratings and narrative comments on the Swiss PRWs.

**Methods:** In November 2017, a random stratified sample of 966 physicians was generated from the regions of Zürich and Geneva. Every selected physician was searched for on 4 rating websites (OkDoc, DocApp, Medicosearch, and Google) between November 2017 and July 2018. It was recorded whether the physician could be identified, what the physician's quantitative rating was, and whether the physician had received narrative comments. In addition, Alexa Internet was used to examine the number of visitors to the PRWs, compared with other websites.

**Results:** Overall, the portion of physicians able to be identified on the PRWs ranged from 42.4% (410/966) on OkDoc to 87.3% (843/966) on DocApp. Of the identifiable physicians, only a few of the selected physicians had been rated quantitatively (4.5% [38/843] on DocApp to 49.8% [273/548] on Google) or received narrative comments (4.5% [38/843] on DocApp to 31.2% [171/548] on Google) at least once. Rated physicians also had, on average, a low number of quantitative ratings (1.47 ratings on OkDoc to 3.74 rating on Google) and narrative comments (1.23 comment on OkDoc to 3.03 comments on Google). All 3 websites allowing ratings used the same rating scale (1-5 stars) and had a very positive average rating: DocApp (4.71), Medicosearch (4.69), and Google (4.41). There were significant differences among the PRWs (with the majority of ratings being posted on Google in past 2 years) and regions (with physicians in Zurich more likely to have been rated and have more ratings on average). Only Google (position 1) and Medicosearch (position 8358) are placed among the top 10,000 visited websites in Switzerland.

**Conclusions:** It appears that this is the first time Google has been included in a study examining physician ratings internationally and it is noticeable how Google has had substantially more ratings than the 3 dedicated PRWs in Switzerland over the past 2 and a half years. Overall, this study indicates that Swiss PRWs are not yet a reliable source of unbiased information regarding patient experiences and satisfaction with Swiss physicians; many selected physicians were unable to be identified, only a few physicians had been rated, and the ratings posted were overwhelmingly positive.

***(J Med Internet Res 2019;21(7):e13816)*** doi:10.2196/13816

**KEYWORDS**

physician rating websites; patient satisfaction; Switzerland

## Introduction

Physician rating websites (PRWs) have been developed in many countries as part of a wider move toward transparency around health care quality, and these allow patients to anonymously rate, comment, and discuss physicians' quality on the Web [1]. Switzerland has been no exception, with the first PRWs in Switzerland, OkDoc and Medicosearch, being established in 2008, at the same time as many international PRWs.

One of the key goals of PRWs is to improve patient welfare through influencing patient decision making by increasing the chance that patients will choose better quality physicians and benefit from this [2] and driving quality improvement by identifying aspects of care needing improvement, so that changes can be made in practice [2]. A related goal of PRWs is to improve patient health literacy to encourage and, therefore, respect patient autonomy [2]. Although recent research suggests that PRWs can influence patient decision making and have an impact on quality improvement [3,4], the ability of PRWs to achieve these goals is likely limited without sufficient number of ratings, as without enough ratings the resulting information is unlikely to be fair for the rated physicians or useful to users of PRWs [5-7].

A recent systematic search of PRWs internationally found that the majority of PRWs were registered in the United States and Germany [8], and the majority of the previous research on PRWs came from these 2 countries. This research has highlighted some key shortcomings of PRWs, including incomplete lists of physicians, low number of physicians rated, and low number of ratings per physician that are overwhelmingly positive. In the United States, Black et al reported in 2009 that their analysis of 6703 ratings of 6101 providers on the PRW RateMDs found that the average number of ratings per physician was 2.7, and their analyses of narrative comments found more positive than negative terms [9]. In 2010, Lagu et al reported that the portion of physicians from a sample of 300 Boston-based physicians that could be identified on 33 websites ranged from 0% to 90.7%; 27% of the sample had been rated once, the average number of ratings per physician was 1.4, and in total there were 190 reviews—170 reviews included quantitative ratings (88% positive), and 66 reviews included narrative comments (89% positive) [5]. In 2011, Kadry et al reported that their analysis of 4999 ratings on PRWs found that physician ratings were mostly positive on all the different rating scales used by PRWs (average of 77 on a 100-point scale, 3.84 on a 5-point scale, and 3.1 on a 4-point scale) [10]. In 2012, Gao et al reported that 16% of the physicians were rated on the PRW RateMDs between 2005 and 2010 [11], the average number of ratings per physician was 3.2, and the ratings were generally positive (mean 3.93 on a 5-point scale). In 2013, Ellimoottil et al reported that in a random sample of 500 urologists [12], 79.6% the of physicians were rated at least once on 10 websites, the average number of reviews per physician was 2.4, 86% of the physicians had a positive rating, and 45% of the physicians had a narrative comment (75% of which were very positive, positive, or neutral). In 2014, Sobin and Goyal reported that in their sample of 281 otolaryngologists, 94.7% could be identified on Healthgrades and 87.9% could be identified on Vitals; of those who were identifiable, 69.9% had been rated at least once on Healthgrades and 81.8% on Vitals, and the average rating was 4.4 on Healthgrades (5-point scale) and 3.4 on Vitals (4-point scale) [13]. In 2017, Murphy et al reported that their analysis of the impact of physician probation on ratings found that average number of ratings per physician was 5.2 for physicians on probation and 4 for controls on Vitals, Healthgrades, and RateMDs, and the average rating on a 5-point scale for physicians on probation was 3.7 compared with 4.0 for controls [14].

In Germany, Strech and Reimann reported in 2012 that from a sample of 298 physicians from Hamburg and Thuringia, 75% to 98% of the physicians could be identified on 6 PRWs, 3% to 26% of the physicians had been rated at least once, the average number of ratings per physician ranged from 1.1 to 3.1, and the average converted standardized rating (1=positive, 2=neutral, and 3=negative) ranged from 1.1 to 1.5 [15]. In 2013, Emmert et al reported that their analysis of 127,192 ratings from 2012 on the German PRW jameda found that 37% of the physicians had been rated, rated physicians had an average of 2.37 ratings, and almost 80% of all ratings were from the 2 best rating categories [16]. In 2014, Emmert et al reported that in their sample of 106 physicians, 96% could be identified on 5 PRWs, 50% of the physicians had been rated at least once, there was an average of 3.08 ratings per physician, and 86% of the ratings were positive (with 75% assigned to the best rating category and only 5% to the worst category) [17]. In 2017, McLennan et al reported that their update study using a sample of 298 physicians from Hamburg and Thuringia found that 65.1% to 94.6% of the physicians could be identified on 6 PRWs, 16% to 83% of the sample had been rated at least once, the average number of ratings per physician ranged from 1.2 to 7.5, and the average converted standardized rating (1=positive, 2=neutral, and 3=negative) ranged from 1.0 to 1.2 [18].

In recent years, there has also been an increasing number of studies published regarding PRWs in China [19-24]. Regarding the frequency of ratings, Hao reported in 2015 that an analysis of the PRW Good Doctor found that 112,873 physicians had received 731,543 quantitative and 772,979 qualitative reviews, on average 37% of the physicians had been reviewed, and the majority of the quantitative reviews were positive (88% positive for treatment effect and 91% positive for bedside manner) [23]. There have also been studies examining the frequency of ratings in other countries. For instance, Liu et al reported in 2018 that their analysis of 640,603 ratings for 57,412 Canadian physicians found that the average number of ratings per physician was 11.2, and the ratings were generally positive with an average of 3.9 (5-point scale) [25]. In 2012, Greaves et al also reported that their analysis of ratings of family practices posted on National Health Service Choices website in the United Kingdom found that 61% of the practices had been rated, and the average number of ratings per practice was 2.1 [26].

Switzerland is a Central European country with a population of about 8.4 million people and 4 official languages (German, French, Italian, and Romansh). The Swiss health care system is highly complex and decentralized, organized around 3 levels of Swiss government (the federal, the cantonal, and the municipalities) [27,28]. All Swiss residents are required to purchase basic mandatory health insurance that is offered by competing nonprofit insurers. Mandatory health insurance covers most general practitioner (GP) and specialist services (among other things), and people not enrolled in managed care plans generally have free choice of professionals. In addition, for-profit insurers offer private complementary insurance for services not covered by mandatory health insurance. Ambulatory physicians (including GPs and specialists) are typically reimbursed in accordance with a standardized fee schedule known as TARMED [27,28].

Although the first PRWs in Switzerland were launched in 2008, there has been limited research conducted on PRWs in Switzerland to date [8,29,30]. Swiss PRWs, however, operate in a rather unique regulatory environment. Owing to Switzerland's restrictive legal framework for data protection, a federal data commissioner decided that negative comments had to be removed from OkDoc, which now only acts as a recommendation portal and explicitly states that any negative comment will be deleted ("Only positive comments recommending your doctor will be accepted. Any negative post will be deleted. Thank you for respecting okdoc's principles!" [author translation]). Although Medicosearch allows negative comments, it informs the concerned physician before publishing it on the Web, so that the physician can decide if the negative feedback is activated. However, if the physician refuses, the feedback function is deactivated, removing also the positive comments [31]. This situation is in stark contrast to more liberal systems (eg, the Federal Court of Justice of Germany confirmed in 2014 the permissibility of ratings on the basis of the right to freedom of expression [32-34]) and likely has important implications in relation to the frequency of ratings and how negative comments are handled on Swiss PRWs.

This study, therefore, examined the frequency of quantitative ratings and narrative comments on Swiss PRWs. In particular, it aimed to explore (1) the number of identifiable physicians on Swiss PRWs, (2) the proportion of physicians with ratings or comments on Swiss PRWs, (3) the average and the maximum number of ratings or comments per physician on Swiss PRWs, (4) the average rating on Swiss PRWs, (5) the website visitor ranking positions of Swiss PRWs, and (6) provide baseline results for future research to assess the development of Swiss PRWs. It is important to examine these issues to help inform future research and health policy in Switzerland in relation to PRWs.

## Methods

### Sample

A random stratified sample of 966 physicians was generated from the regions of Zürich and Geneva. Zürich is the largest city in Switzerland with a total population of 402,762 (December 2016) [35] and is located in north-central Switzerland. Geneva is the second largest city in Switzerland with a total population of 198,979 (December 2016) [35] and is located in south-western Switzerland. The regions of Zürich and Geneva were chosen because of language (German vs French) and comparable number of total physicians (Zürich 3254 physicians and Geneva 2780 physicians) considerations.

In November 2017, all physicians in these regions, working in general practice, obstetrics and gynecology, pediatrics, and dermatology and venereology, were searched for on the Swiss Medical Association's (FMH) medical registry (Ärzteverzeichnis). Specialties were primarily selected based on previous research [15,18]. From each region, a random sample was generated for each specialty based on a 95% confidence level and 5% confidence interval. From Zürich, the random sample comprised 254 of 747 general practice physicians, 85 of 109 obstetrics and gynecology physicians, 74 of 92 pediatrics physicians, and 53 of 61 dermatology and venereology physicians. Therefore, the Zürich sample of 466 physicians represents 46.2% of a total of 1009 physicians. From Geneva, the random sample comprised 272 of 930 general practice physicians, 86 of 111 obstetrics and gynecology physicians, 96 of 128 pediatrics physicians, and 46 of 52 dermatology and venereology physicians. Therefore, the Geneva sample of 500 physicians represents 40.9% of a total of 1221 physicians (see Table 1).

**Table 1.** Physician samples per region.

| Specialty | Zurich | | Geneva | | Total | |
|---|---|---|---|---|---|---|
| | Total physicians found, N | Physicians selected for sample, n (%) | Total physicians found, N | Physicians selected for sample, n (%) | Total physicians found, N | Physicians selected for sample, n (%) |
| General practitioners | 747 | 254 (34.0) | 930 | 272 (29.2) | 1677 | 526 (31.36) |
| Obstetrics and gynecology | 109 | 85 (77.9) | 111 | 86 (77.5) | 220 | 171 (77.7) |
| Pediatrics | 92 | 74 (80.4) | 128 | 96 (75.0) | 220 | 170 (77.3) |
| Dermatology and venereology | 61 | 53 (86.8) | 52 | 46 (88.5) | 113 | 99 (87.6) |
| Total | 1009 | 466 (46.18) | 1221 | 500 (40.95) | 2230 | 966 (43.32) |

### Data Collection

To identify PRWs on which patients can rate and review physicians in Switzerland, a systematic Web-based search was conducted in June 2016 from a patient´s perspective. A total of 10 key search words (see Table 2) in the German language were identified from previously published studies on PRWs conducted in Germany [15,36,37]. As most internet users use a search engine to find health information [36], the systematic search was conducted on Google, which is the most visited search engine in Switzerland, with a market share reported to be 93.5% (Alexa data valid of May 24, 2016). Each search term was searched for and the first 50 hits (5 pages) were examined. As 70% of users only look at the first 2 result pages or less [36], this approach reflects the search behavior of most users. A total of 500 hits were, therefore, examined. A website was included if it allowed users to view quantitative ratings and/or narrative comments about Swiss physicians in a structured manner without having to open an account or log onto the website. Websites that were not dedicated to Swiss physicians were excluded. A total of 3 PRWs were included: OkDoc (found by 8/10 of the search terms), DocApp (found by 4/10 of the search terms), and Medicosearch (found by 2/10 of the search terms). In addition, Google itself allows users to rate and comment on physicians via Google reviews. Furthermore, although the health care information portal doktor does not provide the option for ratings, it links to Google reviews. Google was, therefore, also included in the study, and as far as the author is aware, this is the first time Google has been included in a study examining physician ratings internationally.

Selected physicians were, therefore, searched for on a total of 4 websites: OkDoc, DocApp, Medicosearch, and Google. On each website, every selected physician was searched for between November 2017 and July 2018, and it was recorded in a SPSS file whether the physician could be found, the physician's rating, the number of ratings and narrative comments, and the text of narrative comments. As OkDoc now only allows recommendations, the number of these recommendations were assigned to the *number of ratings*. All websites allowing ratings (DocApp, Medicosearch, and Google) used the same rating scale (1-5 stars); a rating of 4 to 5 stars was considered a positive rating, 3 stars a neutral rating, and 1 to 2 stars a negative rating.

Alexa Internet was used to examine the number of visitors to PRWs, compared with other websites. Founded in 1996, Alexa provides commercial Web traffic data and analytics. Traffic estimates are based on data from a global traffic panel and from websites that have chosen to install the Alexa script on their site and certify their metrics. The Alexa global traffic ranking is based on the estimated average of daily unique visitors and its estimated number of page views over the past 3 months relative to all other websites. In addition, Alexa provides a similar country-specific ranking based on how a website ranks relative to other websites in a particular country over the past month. The PRWs were searched for on Alexa in November 2017 and their Switzerland specific ranking was recorded.

**Table 2.** Systematic search of the Swiss physician rating websites on Google.

| Number | Search terms in German | English translation | Physician rating website found |
|---|---|---|---|
| 1 | Arztsuche | Physician search | None |
| 2 | Arzt finden | Find a physician | OkDoc, DocApp |
| 3 | Arzt bewerten | Rate my physician | OkDoc, Medicosearch, DocApp |
| 4 | Arztbewertung | Physician rating | OkDoc, DocApp |
| 5 | Arzt empfehlen | Recommend a physician | OkDoc |
| 6 | Arztempfehlung | Physician recommendation | None |
| 7 | Ärzte Beurteilungen | Physician reviews | OkDoc |
| 8 | Online Arztbewertung | Online physician rating | OkDoc, DocApp, Medicosearch |
| 9 | Arztbewertungsportal | Physician rating website | OkDoc |
| 10 | Guter Arzt | Good physician | OkDoc |

### Data Analysis

Descriptive statistics included means and standard deviations for continuous variables and percentages for categorical variables. To analyze whether differences exist within an individual PRW as well as across PRWs between the 2 regions (Zürich and Geneva) and between GPs and specialists (obstetrics and gynecology, pediatrics, and dermatology and venereology), chi-squared tests were used for categorical data and *t* tests for continuously distributed data. To analyze differences across the PRWs, a *sum score* was created; for example, in relation to how many physicians were identified on PRWs, a score ranging from 0 (not identified on any PRW) to 4 (identified on all PRWs) was created and, subsequently, it was analyzed whether the mean of this score was different between the 2 groups being examined. All analyses were performed with a significance level alpha set to .05 and 2-tailed tests, using Statistical Package for the Social Sciences (SPSS version 24 for Windows, IBM Corporation).

## Results

The full results regarding the quantitative ratings and narrative comments are presented in Tables 3-5. See Multimedia Appendix 1 for the full results of comparisons between the 2 regions and Multimedia Appendix 2 for the results of comparisons between GPs and specialists.

**Table 3.** Quantitative rating.

| Region, website | Physicians, n (%) | | Ratings per physician | | Rating, mean (SD) |
|---|---|---|---|---|---|
| | Identifiable | Rated[a] | Mean (SD) | Maximum | |
| **Zurich (N=466)** | | | | | |
| OkDoc | 225 (48.3) | 35 (15.5) | 1.26 (0.6) | 4 | —[b] |
| DocApp | 406 (87.1) | 37 (9.1) | 2.38 (5.2) | 32 | 4.70 (0.7) |
| Medicosearch | 356 (76.4) | 74 (20.7) | 2.78 (5.3) | 32 | 4.68 (0.8) |
| Google | 268 (57.5) | 150 (55.9) | 4.56 (5.9) | 56 | 4.38 (0.9) |
| **Geneva (N=500)** | | | | | |
| OkDoc | 185 (37.0) | 41 (22.1) | 1.66 (1.1) | 6 | — |
| DocApp | 437 (87.4) | 1 (0.2) | 1 | 1 | 5 |
| Medicosearch | 331 (66.2) | 22 (6.6) | 1.23 (0.5) | 3 | 4.73 (0.8) |
| Google | 280 (56.0) | 123 (43.9) | 2.67 (2.2) | 13 | 4.45 (0.9) |
| **Overall (N=966)** | | | | | |
| OkDoc | 410 (42.4) | 76 (18.5) | 1.47 (0.9) | 6 | — |
| DocApp | 843 (87.3) | 38 (4.5) | 2.34 (5.1) | 32 | 4.71 (0.7) |
| Medicosearch | 687 (71.1) | 96 (13.9) | 2.42 (4.7) | 32 | 4.69 (0.8) |
| Google | 548 (56.7) | 273 (49.8) | 3.74 (4.7) | 56 | 4.41 (0.9) |

[a]Each n value is a sample from the identifiable physician population value.
[b]Data not applicable.

**Table 4.** Narrative comments.

| Region, website | Physicians, n (%) | | Comments per physician | |
|---|---|---|---|---|
| | Identifiable | With comments | Mean (SD) | Maximum |
| **Zurich (N=466)** | | | | |
| OkDoc | 225 (48.3) | 18 (8.0) | 1.11 (0.3) | 2 |
| DocApp | 406 (87.1) | 37 (9.1) | 2.38 (5.2) | 32 |
| Medicosearch | 356 (76.4) | 74 (20.7) | 2.78 (5.3) | 32 |
| Google | 268 (57.5) | 104 (38.8) | 3.7 (5.5) | 49 |
| **Geneva (N=500)** | | | | |
| OkDoc | 185 (37.0) | 13 (7.0) | 1.38 (0.7) | 3 |
| DocApp | 437 (87.4) | 1 (0.2) | 1 | 1 |
| Medicosearch | 331 (66.2) | 22 (6.6) | 1.27 (0.6) | 3 |
| Google | 280 (56.0) | 67 (23.9) | 2.0 (2.1) | 12 |
| **Overall (N=966)** | | | | |
| OkDoc | 410 (42.4) | 31 (7.5) | 1.23 (0.5) | 3 |
| DocApp | 843 (87.3) | 38 (4.5) | 2.34 (5.1) | 32 |
| Medicosearch | 687 (71.1) | 96 (13.9) | 2.44 (4.7) | 32 |
| Google | 548 (56.7) | 171 (31.2) | 3.04 (4.6) | 49 |

**Table 5.** Distribution of narrative comments.

| Region, website | Total, N | Distribution of comments, n (%) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2018 (half year) | 2017 | 2016 | 2015 | 2014 | 2013 | 2012 | 2011 | 2010 | 2009 | 2008 |
| **Zurich** | | | | | | | | | | | | |
| OkDoc | 20 | 0 | 0 | 1 (5) | 0 | 0 | 0 | 0 | 2 (10) | 2 (10) | 3 (15) | 12 (60) |
| DocApp | 56 | 3 (5) | 22 (39) | 24 (43) | 7 (13) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medicosearch | 206 | 6 (2.9) | 57 (27.6) | 59 (28.6) | 6 (2.9) | 8 (3.8) | 11 (5.3) | 16 (7.8) | 18 (8.7) | 12 (4.9) | 12 (5.8) | 1 (0.5) |
| Google | 386 | 160 (41.4) | 187 (48.4) | 29 (7.5) | 5 (1.2) | 1 (0.2) | 3 (0.7) | 0 | 0 | 1 (0.2) | 0 | 0 |
| Total | 668 | 169 (25.2) | 266 (39.8) | 113 (16.9) | 18 (2.6) | 9 (1.3) | 14 (2.0) | 16 (2.3) | 20 (2.9) | 15 (2.2) | 15 (2.2) | 13 (1.9) |
| **Geneva** | | | | | | | | | | | | |
| OkDoc | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 (6) | 0 | 3 (17) | 14 (78) |
| DocApp | 1 | 0 | 1 (100) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medicosearch | 28 | 0 | 0 | 2 (7) | 0 | 2 (7) | 2 (7) | 1 (4) | 5 (18) | 7 (25) | 9 (32) | 0 |
| Google | 134 | 75 (55.9) | 39 (29.1) | 12 (8.9) | 2 (1.5) | 4 (2.9) | 2 (1.5) | 0 | 0 | 0 | 0 | 0 |
| Total | 181 | 75 (41.4) | 40 (22.0) | 14 (7.7) | 2 (1.1) | 6 (3.3) | 4 (2.2) | 1 (0.5) | 6 (3.3) | 7 (3.8) | 12 (6.6) | 14 (7.7) |
| **Overall** | | | | | | | | | | | | |
| OkDoc | 38 | 0 | 0 | 1 (3) | 0 | 0 | 0 | 0 | 3 (8) | 2 (5) | 6 (16) | 26 (68) |
| DocApp | 57 | 3 (5) | 23 (40) | 24 (42) | 7 (12) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medicosearch | 234 | 6 (2.5) | 57 (24.3) | 61 (26) | 6 (2.5) | 10 (4.2) | 13 (5.5) | 17 (7.2) | 23 (9.8) | 19 (8.1) | 21 (8.9) | 1 (0.4) |
| Google | 520 | 235 (45.1) | 226 (43.4) | 41 (7.8) | 7 (1.3) | 5 (0.1) | 5 (0.1) | 0 | 0 | 1 (0.1) | 0 | 0 |
| Total | 849 | 244 (28.7) | 306 (36) | 127 (14.9) | 20 (2.3) | 15 (1.7) | 18 (2.1) | 17 (2.0) | 26 (3.0) | 22 (2.5) | 27 (3.1) | 27 (3.1) |

## Quantitative Ratings

### Identifiable Physicians

Overall, the portion of physicians from the random sample that were able to be identified on the selected PRWs ranged from 42.4% (410/966) on OkDoc to 87.3% (843/966) on DocApp. Physicians were identified significantly more in Zurich on OkDoc ($X^2_1$=12.6; *P*<.001) and Medicosearch ($X^2_1$=12.2, *P*<.001). Across all PRWs, there was also a significant difference between Zurich (mean 2.7; SD 1.2) and Geneva (mean 2.5; SD 1.2) ($t_{964}$=2.9; *P*=.004). GPs were identified significantly more than specialists on Medicosearch ($X^2_1$=7.3; *P*=.007).

### Rated Physicians

Overall, of the physicians identified, the portion that had been rated at least once ranged from 4.5% (38/843) on DocApp to 49.8% (273/548) on Google. Physicians from Zurich were rated significantly more on DocApp ($X^2_1$=38.6; *P*<.001), Medicosearch ($X^2_1$=28.5; *P*<.001), and Google ($X^2_1$=7.9; *P*=.005). Across all PRWs, there was also a significant difference between Zurich (mean 1.0; SD 0.9) and Geneva (mean 0.8; SD 0.8) ($t_{274}$=2.0; *P*=.046). GPs were rated significantly on Google ($X^2_1$=19.1; *P*<.001). Across all PRWs, there was also a significant difference between GPs (mean 0.9; SD 0.8) and specialists (mean 1.1; SD 0.9) ($t_{274}$=−2.2; *P*=.03).

### Average and Maximum Number of Ratings

Overall, the average number of ratings per physician ranged from 1.47 (SD 0.9) on OkDoc to 3.74 (SD 4.7) on Google. The maximum number of ratings per physician ranged from 6 on OkDoc to 56 on Google. Whereas the physicians in Geneva (mean 1.6; SD 1.1) had significantly more ratings on average than the physicians in Zurich (mean 1.3; SD 0.6) on OkDoc ($t_{65}$=2.1; *P*=.04), the physicians in Zurich had significantly more ratings on average than the physicians in Geneva on Medicosearch (mean 2.8, SD 5.3 vs mean 1.2, SD 0.5; $t_{77}$=2.5; *P*=.02) and Google (mean 4.6, SD 5.9 vs mean 2.7, SD 2.2; $t_{198}$=3.7; *P*<.001). Similarly, whereas GPs (mean 1.7; SD 1.1) had significantly more ratings on average than specialists (mean 1.3; SD 0.6) on OkDoc ($t_{57}$=2.1; *P*=.04), specialists (mean 4.5; SD 5.5) had significantly more ratings on average than GPs (mean 2.8; SD 3.2) on Google ($t_{249}$=−3.2; *P*=.001).

### Average Rating

Overall, the 3 websites allowing ratings all used the same rating scale (1-5 stars) and had a very positive average rating: DocApp, 4.71; Medicosearch, 4.69; and Google, 4.41. There were no significant differences between the regions or between GPs and specialists.

## Narrative Comments

### *Physicians With Comments*

Overall, of the physicians identified, the portion that had received at least 1 comment ranged from 4.5% (38/843) on DocApp to 31.2% (171/548) on Google. Physicians from Zurich had received a comment significantly more often than Geneva physicians on DocApp ($X^2_1$=38.2; *P*<.001) and Google ($X^2_1$=14.8; *P*<.001). GPs also had received a comment significantly more often than specialists on Google ($X^2_1$=23.1; *P*<.001).

### *Average and Maximum Number of Comments*

Overall, the average number of comments per physician ranged from 1.23 (SD 0.5) on OkDoc to 3.04 (SD 4.6) on Google. The maximum number of comments per physician ranged from 3 on OkDoc to 49 on Google. Physicians from Zurich had significantly more comments on average than physicians in Geneva on Medicosearch (mean 2.8, SD 5.3 vs mean 1.3, SD 0.6; $t_{77}$=2.4; *P*=.02) and Google (mean 3.7, SD 5.6 vs mean 2.0, SD 2.1; $t_{142}$=2.9; *P*=.005). There were no significant differences between GPs and specialists.

### *Distribution of Comments*

Overall, the selected physicians in the sample had a total number of 849 comments from 2008 to 2018 (half year), with 80% of comments (677/849) having been posted during the last 2 and a half years (2016 to 2018). The majority of comments in Zurich (386/668, 57.7%) and Geneva (134/181, 74.0%) were made on Google. OkDoc only had 1 comment posted for all 966 physicians in the sample during the last 5 and a half years (2012-2018). Physicians in the Zurich sample also had substantially more comments (668 comments) compared with physicians in the Geneva sample (181 comments), with 78.7% (668/849) of total comments coming from physicians in Zurich (see Figures 1-3).

**Figure 1.** Distribution of comments in Zurich.

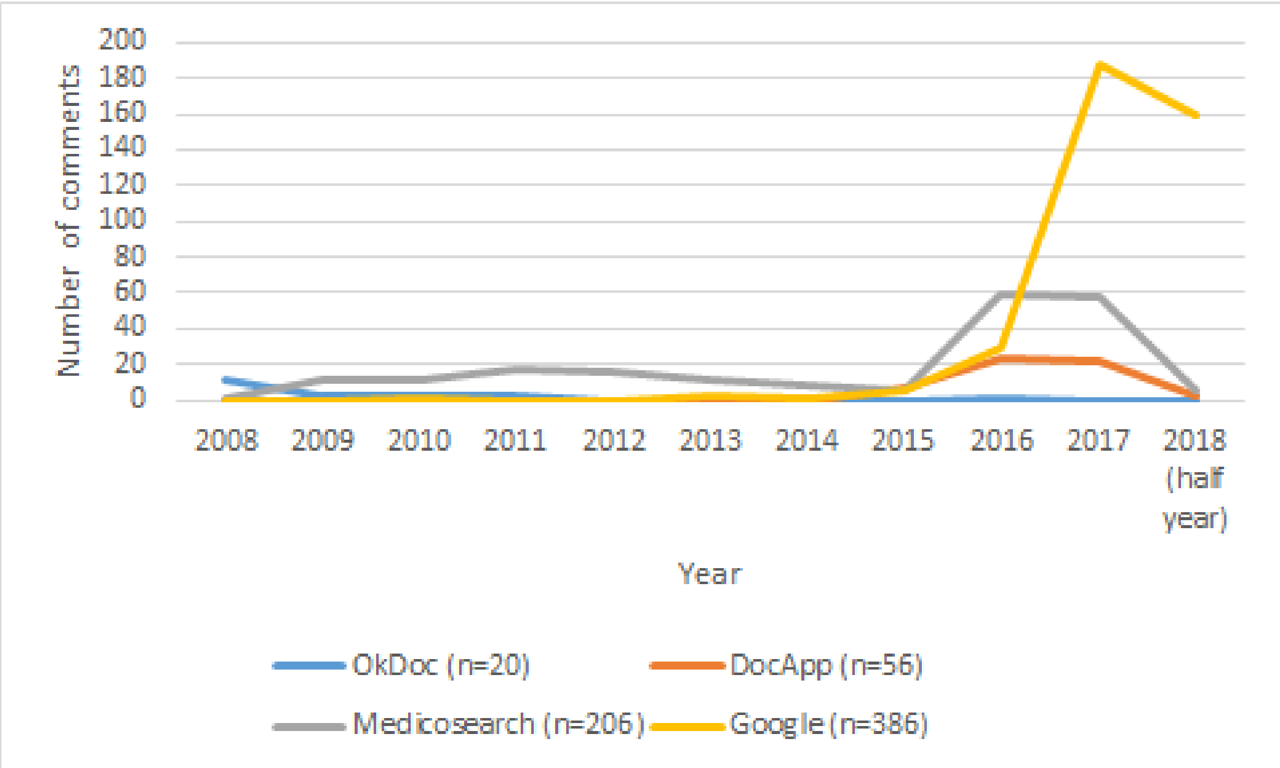


**Figure 2.** Distribution of comments in Geneva.

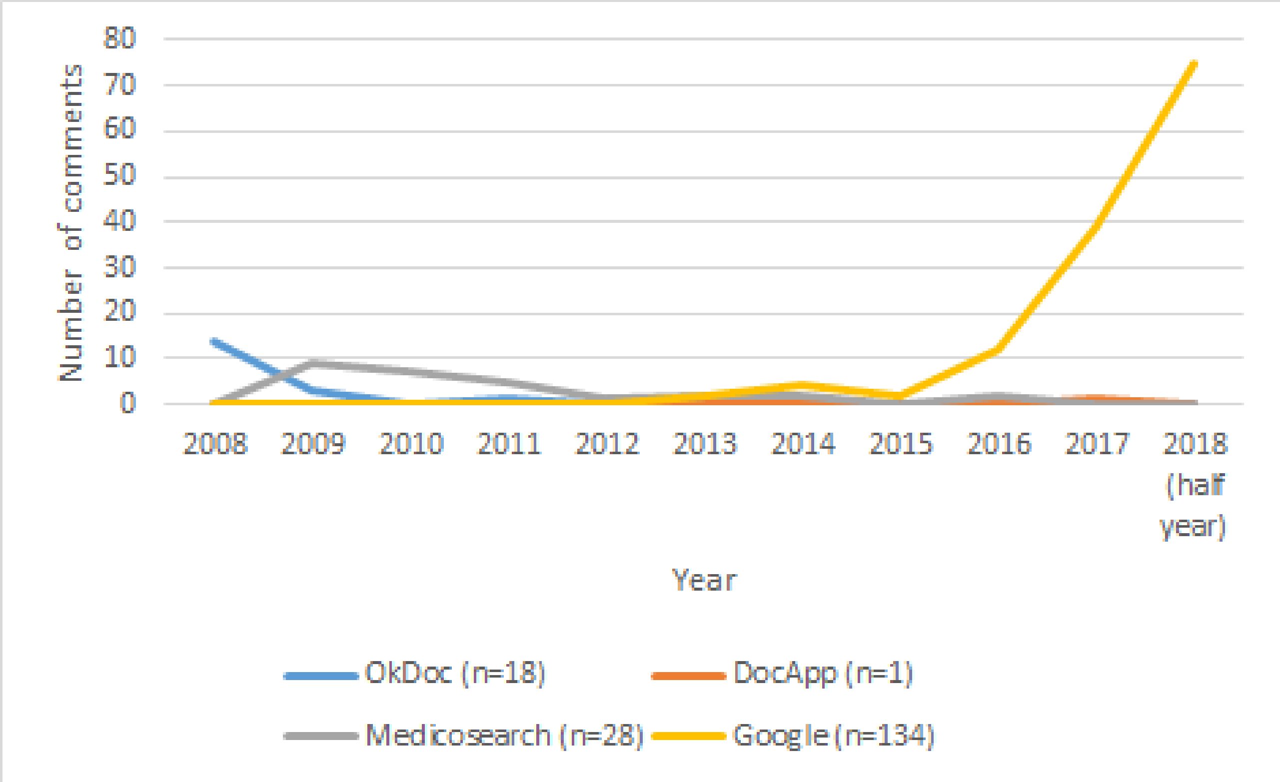
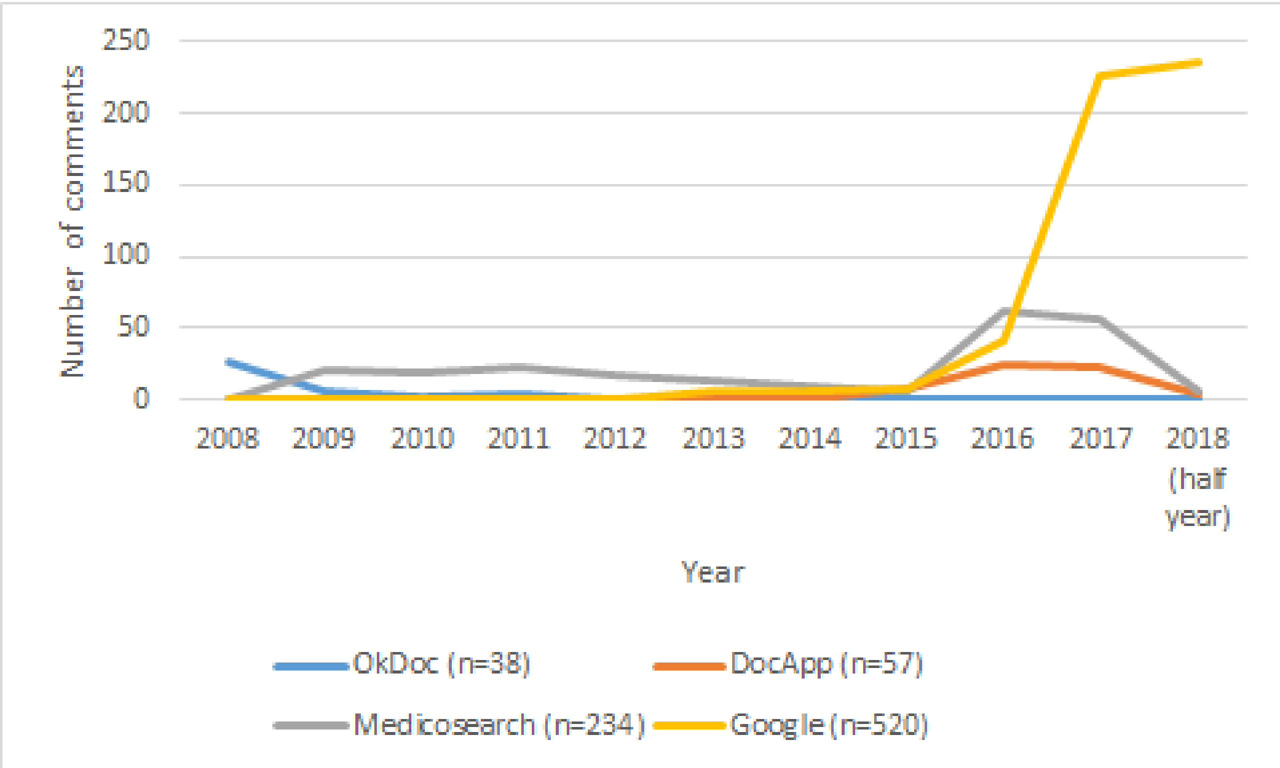



**Figure 3.** Distribution of comments in overall.




## Website Visitor Ranking Positions

Whereas Google was in position 1 on Alexa for the most visited websites in Switzerland, Alexa indicated that the use of the dedicated PRWs was not common, with only Medicosearch (position 8358) placed among the top 10,000 visited websites in Switzerland. DocApp was ranked 19,858, whereas there were insufficient data for OkDoc. In comparison, the hotel rating site TripAdvisor ranked 154. Rankings are as of November 10, 2017.

## Discussion

### Principal Findings

This is the first study to examine the frequency of ratings on Swiss PRWs and it has resulted in a number of key findings: (1) many of the selected physicians could not be identified on Swiss PRWs, (2) very few of the selected physicians had been rated quantitatively or qualitatively and those who had been rated had on average a low number of ratings, (3) there were significant differences among the PRWs, with Google having substantially more ratings in the past 2 and a half years than the 3 dedicated PRWs, (4) there were also significant differences between regions, with physicians in Zurich more likely to have been rated and have more ratings on average, and (5) all 3 websites allowing ratings had a very positive average rating overall.

#### *Identifiable Physicians*

Incomplete lists of physicians have been identified as a weakness of many PRWs internationally [5,6], and it appears that the users of PRWs in Switzerland are also not able to find many physicians on Swiss PRWs, with the overall portion of selected physicians that was able to be identified ranging from 42.4% on OkDoc to 87.3% on DocApp. Although the result for OkDoc may not be overly surprising, as it appears to be rather inactive following the decision of a federal data commissioner that negative comments have to be removed [31], only 56.7% of the physicians could be identified on Google. This range is lower compared with the findings of studies on PRWs in other countries. For instance, a 2010 study in Germany found that the portion of physicians that could be identified on German PRWs ranged from 76% to 98% [15], whereas a follow-up study in 2014 found a range of 65% to 95% [18]. The current lack of comprehensiveness of Swiss PRWs could inhibit their usefulness, and it will be important to consider how more complete lists of physicians can be provided.

#### *Number of Ratings*

It is widely considered that a key factor in PRWs being successful in their goals of influencing patients' decision making and driving quality improvement is having sufficient number of ratings [5,6]. However, low number of ratings has been identified as a key shortcoming of PRWs in many countries, which has called into question their representativeness, validity, and usefulness [5,6]. This study also indicates that insufficient rating can also be an issue for Swiss PRWs, with only a few of the identifiable physicians having been rated quantitatively (4.5% on DocApp to 49.8% on Google) or qualitatively (4.5% on DocApp to 31.2% on Google) at least once. Rated physicians also had on average a low number of quantitative ratings (1.47 ratings on OkDoc to 3.74 rating on Google) and narrative comments (1.23 comment on OkDoc to 3.03 comments on Google). Although the results of this study were lower than those found in a recent study in Germany, which found that 16% to 83% of the sample had been rated at least once and these physicians had an average number of ratings between 1.2 and 7.5 [18], they were very similar to the results of the previous studies in Germany [15-17] and the United States [5,9,11,12]. However, it should be noted that many of these studies' reported figures were, unlike this study, portions of the total sample rather than a portion of the identifiable physicians and were, therefore, slightly higher than reported.

There is, however, currently limited research internationally examining the reasons why patients do not rate their physicians on PRWs. The use of PRWs first requires the public to be aware of them [38]. Recent studies in the United States and Germany suggest that a lack of awareness is no longer a key barrier to PRW usage in these countries [38-41], although a recent study in the England found public awareness of PRWs to still be very low [42]. However, despite the fact that awareness of PRWs is an important factor, it should be noted that although the studies conducted in the United States and Germany found high awareness of the PRWs, the level of PRW usage was still found to be comparable with previous studies [3,43], suggesting that even if awareness of PRWs increases, there are other factors behind the low level of physician ratings. A recent qualitative study in Germany aiming to examine these other factors identified 2 key overarching groups of factors—first, factors concerning the physician-patient relationship and second, factors regarding the technical aspects of PRWs [44]. Although a qualitative study in 2016 with participants residing in the German-speaking part of Switzerland also highlighted the need to improve the design of PRWs, the study involved German PRWs rather than Swiss PRWs [29]. Further research is, therefore, needed in Switzerland regarding public awareness of PRWs and factors influencing patients' decision to rate or not rate physicians.

There were significant differences among the PRWs in relation to the frequency of ratings. Although OkDoc was the first Swiss PRWs, launched in 2008, it was evident how inactive the website had become since the federal data commissioner decided that negative comments had to be removed. Although the portion of the rated physicians on OkDoc is still higher than DocApp and Medicosearch, the distribution of comments indicates that this is because of ratings posted in 2008. Indeed, OkDoc only had 1 comment posted for all 966 physicians in the sample during the last 5 and a half years (2012-2018). This situation is possibly exacerbated by Okdoc being the only Swiss PRW not to offer an English version of the website, given the high number of foreign residents in Switzerland [45].

In contrast, it is noticeable how Google has had substantially more quantitative ratings and narrative comments than the 3 dedicated PRWs in the past 2 and a half years, and how it has been able to establish itself as the most used website in Switzerland for physician ratings. It remains to be seen whether the other dedicated PRWs will be able to increase their number of physician ratings in the future or whether Google will continue to dominate the market. Future updates will be helpful to assess how this develops. In the meantime, given the current large differences among the PRWs in terms of how many physicians can be identified and the number physician ratings, it would be advisable for the users of PRWs to utilize a number of PRWs when searching for a new physician.

There were also significant differences between the 2 regions (Zurich and Geneva) in relation to the frequency of ratings, with physicians from Zurich having been rated at least once more often and having on average more ratings. It is, however, unclear what the reason is behind these differences between Zurich and Geneva. These differences may simply reflect differences in the networks of the PRWs or may be a result of more cultural factors. Previous research in Switzerland has indicated that the German-speaking Swiss tend to be more critical toward their physicians and less-dependent on them, compared with the French- and the Italian-speaking Swiss [46,47]. However, further research is needed to examine what is causing these differences in the use of Swiss PRWs.

### *Average Rating*

Although there have been concerns from the medical profession that PRWs would be primarily used for *doctor-bashing* [48,49], these fears have proved to be unfounded with the previous international research finding ratings on PRWs to be on average very positive [5,10-12,16-18]. This study has found a similar situation in Switzerland; the 3 PRWs that allow ratings all used the same rating scale (1-5 stars) and had a very positive average rating: DocApp, 4.71; Medicosearch, 4.69; and Google, 4.41. Such overwhelmingly positive ratings also raise concerns about the representativeness, validity, and usefulness of information on PRWs [5,6].

In Switzerland, it appears that the restrictive legal framework regarding data protection may be having a huge impact on the types of ratings that are on Swiss PRWs. As a key goal of PRWs is to promote transparency, this is concerning and suggests that Swiss PRWs are not a reliable source of unbiased information regarding patient experiences and satisfaction with Swiss physicians. Addressing the potential harms to physicians without limiting the potential health literacy benefits for patients is challenging; however, as Strech has noted: “In many countries the medical profession enjoys privileges such as strong advocacy groups and special social facilities. Thus, the denial of transparency on patient experiences and satisfaction (with physician performance) requires a strong rationale” [2]. Further consideration is needed to determine whether the current lack of transparency on Swiss PRWs is justified or whether changes are required.

## Limitations

This study has a number of limitations that should be taken into account when interpreting the results. First, although a systematic Web-based search of Swiss PRWs was conducted, there might be other types of websites that allow Swiss physicians to be rated, which were not included in this study. This is a fast-moving area and it does appear that there are some websites that have started allowing ratings or making ratings publicly available after this project commenced (eg, deindoktor and doctena), which should be added to any future studies examining PRWs in Switzerland. Second, only German search terms were used for the systematic Web-based search of Swiss PRWs. Although the author is confident that no important Swiss PRWs were missed at the time of developing and conducting the project, it would be preferable if French and Italian search terms are also included in future research in Switzerland to ensure that no PRWs are being missed. Third, the sample was only taken from 2 regions in Switzerland, which might have limited the generalizability of the results. Although the study used a representative random sample from a German- and a French-speaking region of Switzerland with comparable number of physicians, given the significant differences found between the 2 regions, it would be helpful for future research to include other regions to examine whether these differences can be found between other German- and French-speaking regions and in the Italian-speaking region of Ticino. Finally, because of practical considerations of searching for 966 physicians on 4 different websites, data were collected over a 9-month period. This might have had led to differences among the PRWs that were examined at the beginning of the data collection compared with those PRWs examined at the end of data collection.

## Conclusions

With a growing number of patients utilizing the internet in relation to their health care [50], it is expected that PRWs will play an increasingly important role in selecting a new physician. However, for PRWs to be helpful for the users, and fair for the rated physicians, it is important that PRWs have a sufficient number of ratings. This study indicates that Swiss PRWs are currently not an effective mechanism of collecting patient experiences as a source of information for others; many physicians could not be identified, of the physicians identified, most had not been rated, and those that had been rated had on average only a few ratings. However, there were significant differences among the PRWs. As far as the author is aware, this is the first time Google has been included in a study examining physician ratings internationally and it is noticeable how Google has had substantially more ratings than the 3 dedicated Swiss PRWs in the past 2 and a half years. This is an important development not previously reported in the context of public reporting activities. Given Google’s general market dominance globally, Google might become the primary website for physician rating and it will be important for health systems to reflect on the implications of this. However, in the meantime, given the current large differences among the PRWs, it would be advisable for the users of PRWs to utilize a number of PRWs when searching for a new physician [18]. However, in addition to the low number of ratings, the ratings that are on Swiss PRWs are overwhelmingly positive, which suggests that Swiss PRWs are also not a reliable source of unbiased information regarding patient experiences and satisfaction with Swiss physicians. Although more research is needed to examine the factors influencing the low number of ratings and the lack of negative ratings in Switzerland, it appears that Switzerland’s restrictive legal framework regarding data protection may be a key factor. There is a need for more consideration to be given to the correct equilibrium between protecting physicians from harm and promoting patients’ autonomy and health literacy. However, as long as the current restrictive legal framework remains, the utility of Swiss PRWs is likely to be weakened from the patient’s point of view. Swiss PRWs should seek to enrich the utility of their websites with additional features, such as the possibility to book an appointment with a physician through the PRW (as already offered by Medicosearch and other websites, such as deindoktor and doktena) and providing information about the language spoken by the specific physician (as already offered by DocApp and Medicosearch). Such features may increase the utility of the PRWs and perhaps also help increase the number of physician ratings in the long term (eg, by sending invites to patients after the appointments they have booked on the Web).

## Acknowledgments

The author would like to thank Dr Michael Rost for his assistance with running the statistical tests and to the anonymous peer reviewers for their helpful comments. This study was funded by the Swiss Academy of Medical Sciences' Käthe-Zingg-Schwichtenberg-Fonds, which had no role in the study design, collection, analysis or interpretation of data, writing of the paper, or decision to submit the paper for publication.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Results of comparisons between regions.

[PDF File (Adobe PDF File), 106KB - jmir_v21i7e13816_app1.pdf ]

## Multimedia Appendix 2

Results of comparisons between specialities.

[PDF File (Adobe PDF File), 106KB - jmir_v21i7e13816_app2.pdf ]

## References

1. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 01;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
2. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011 Dec 06;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]
3. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
4. Emmert M, Meszmer N, Sander U. Do health care providers use online patient ratings to improve the quality of care? Results from an online-based cross-sectional study. J Med Internet Res 2016 Dec 19;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]
5. Lagu T, Hannon N, Rothberg M, Lindenauer P. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
6. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
7. Patel S, Cain R, Neailey K, Hooberman L. General practitioners' concerns about online patient feedback: findings from a descriptive exploratory qualitative study in England. J Med Internet Res 2015 Dec 08;17(12):e276 [FREE Full text] [doi: 10.2196/jmir.4989] [Medline: 26681299]
8. Rothenfluh F, Schulz P. Content, quality, and assessment tools of physician-rating websites in 12 countries: quantitative analysis. J Med Internet Res 2018 Dec 14;20(6):e212 [FREE Full text] [doi: 10.2196/jmir.9105] [Medline: 29903704]
9. Black E, Thompson L, Saliba H, Dawson K, Black N. An analysis of healthcare providers' online ratings. Inform Prim Care 2009;17(4):249-253 [FREE Full text] [Medline: 20359403]
10. Kadry B, Chu L, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
11. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb 24;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
12. Ellimoottil C, Hart A, Greco K, Quek M, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
13. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? JAMA Otolaryngol Head Neck Surg 2014 Jul;140(7):635-638. [doi: 10.1001/jamaoto.2014.818] [Medline: 24876073]

14. Murphy G, Awad M, Osterberg E, Gaither T, Chumnarnsongkhroh T, Washington S, et al. Web-based physician ratings for California physicians on probation. J Med Internet Res 2017 Dec 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
15. Strech D, Reimann S. [German-language doctor rating portals: the status quo of their evaluation criteria, evaluation tendencies and usage]. Gesundheitswesen 2012;74:e61-e67.
16. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 06;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
17. Emmert M, Gerstner B, Sander U, Wambach V. [An inventory of assessments on doctor rating portals using the example of the Nuremberg Health Network Quality and Efficiency]. Gesundh ökon Qual manag 2014;19:167.
18. McLennan S, Strech D, Reimann S. Developments in the frequency of ratings and evaluation tendencies: a review of German physician rating websites. J Med Internet Res 2017 Dec 25;19(8):e299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]
19. Zhang W, Deng Z, Hong Z, Evans R, Ma J, Zhang H. Unhappy patients are not alike: content analysis of the negative comments from China's good doctor website. J Med Internet Res 2018 Dec 25;20(1):e35 [FREE Full text] [doi: 10.2196/jmir.8223] [Medline: 29371176]
20. Li J, Liu M, Li X, Liu X, Liu J. Developing embedded taxonomy and mining patients' interests from web-based physician reviews: mixed-methods approach. J Med Internet Res 2018 Dec 16;20(8):e254 [FREE Full text] [doi: 10.2196/jmir.8868] [Medline: 30115610]
21. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The effect of online effort and reputation of physicians on patients' choice: 3-wave data analysis of China's good doctor website. J Med Internet Res 2019 Mar 08;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]
22. Hao H, Zhang K. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]
23. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
24. Li Y, Ma X, Song J, Yang Y, Ju X. Exploring the effects of online rating and the activeness of physicians on the number of patients in an online health community. Telemed J E Health 2019 Jan 24;20(8):e254. [doi: 10.1089/tmj.2018.0192] [Medline: 30676279]
25. Liu J, Matelski J, Bell C. Scope, breadth, and differences in online physician ratings related to geography, specialty, and year: observational retrospective study. J Med Internet Res 2018 Dec 07;20(3):e76 [FREE Full text] [doi: 10.2196/jmir.7475] [Medline: 29514775]
26. Greaves F, Pape U, Lee H, Smith D, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012 Oct 17;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
27. De Pietro C, Camenzind P, Sturny I, Crivelli L, Edwards-Garavoglia S, Spranger A, et al. Switzerland: health system review. Health Syst Transit 2015;17(4):1-288, xix [FREE Full text] [Medline: 26766626]
28. Sturny I. International Health Care System Profiles. The Swiss Health Care System URL: https://international.commonwealthfund.org/countries/switzerland/ [WebCite Cache ID 784AFrzjb]
29. Rothenfluh F, Germeni E, Schulz P. Consumer decision-making based on review websites: are there differences between choosing a hotel and choosing a physician? J Med Internet Res 2016 Dec 16;18(6):e129 [FREE Full text] [doi: 10.2196/jmir.5580] [Medline: 27311623]
30. Rothenfluh F, Schulz P. Physician rating websites: what aspects are important to identify a good doctor, and are patients capable of assessing them? A mixed-methods approach including physicians' and health care consumers' perspectives. J Med Internet Res 2017 Dec 01;19(5):e127 [FREE Full text] [doi: 10.2196/jmir.6875] [Medline: 28461285]
31. Roos N. Schweizer Radio und Fernsehen. 2014. [Doctors' reviews on the Internet still little meaningful] URL: https://www.srf.ch/sendungen/kassensturz-espresso/themen/gesundheit/aerzte-bewertungen-im-internet-noch-wenig-aussagekraeftig [accessed 2019-02-14] [WebCite Cache ID 76BgL5ONT]
32. [Press Office of the Federal Court of Justice]. [Federal Court of Justice rejects the claim of a doctor to delete his data from a Ärztebewertungsportal] URL: http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&sid=18e188ccd712bed4e8781314de1c4c63&anz=1&pos=0&nr=68915&linked=pm&Blank=1 [accessed 2019-07-15] [WebCite Cache ID 784IzaGoJ]
33. Frankfurter Allgemeine. [Doctors need to be anonymously evaluated on the internet] URL: https://www.faz.net/aktuell/finanzen/meine-finanzen/geld-ausgeben/nachrichten/aerzteportale-jameda-urteil-az-bgh-vi-zr-358-13-13169241.html [accessed 2019-05-02] [WebCite Cache ID 784JPra7k]
34. Zeit Online. [Doctors have to accept reviews on the internet] URL: https://www.zeit.de/digital/datenschutz/2014-09/aerzte-bewertung [accessed 2019-05-02] [WebCite Cache ID 784JcLbCi]

35. [Federal Office for statistics]. [City Statistics: Age Structure] URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/querschnittsthemen/city-statistics/indikatoren-lebensqualitaet/demographischer-kontext/altersstruktur.html [accessed 2019-07-15] [WebCite Cache ID 784K07ZE1]
36. Emmert M, Sander U, Esslinger A, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
37. Emmert M, Maryschok M, Eisenreich S, Schöffski O. [Websites to assess quality of care--appropriate to identify good physicians?]. Gesundheitswesen 2009 Apr;71(4):e18-e27. [doi: 10.1055/s-0028-1103288] [Medline: 19221986]
38. Hibbard J, Berkman N, McCormack L, Jael E. The impact of a CAHPS report on employee knowledge, beliefs, and decisions. Med Care Res Rev 2002 Mar;59(1):104-116. [doi: 10.1177/107755870205900106] [Medline: 11877876]
39. Hanauer D, Zheng K, Singer D, Gebremariam A, Davis M. Parental awareness and use of online physician rating sites. Pediatrics 2014;134:e966-e975 [FREE Full text]
40. Hanauer D, Zheng K, Singer D, Gebremariam A, Davis M. Public awareness, perception, and use of online physician rating sites. J Am Med Assoc 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
41. McLennan S, Strech D, Meyer A, Kahrass H. Public awareness and use of German physician ratings websites: cross-sectional survey of four north German cities. J Med Internet Res 2017 Dec 09;19(11):e387 [FREE Full text] [doi: 10.2196/jmir.7581] [Medline: 29122739]
42. Patel S, Cain R, Neailey K, Hooberman L. Public awareness, usage, and predictors for the use of doctor rating websites: cross-sectional study in England. J Med Internet Res 2018 Dec 25;20(7):e243 [FREE Full text] [doi: 10.2196/jmir.9523] [Medline: 30045831]
43. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar 31;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
44. McLennan S, Strech D, Kahrass H. Why are so few patients rating their physicians on German physician rating websites? A qualitative study. BMC Health Serv Res 2018 Aug 29;18(1):670 [FREE Full text] [doi: 10.1186/s12913-018-3492-0] [Medline: 30157842]
45. Federal Statistical Office. Foreign population URL: https://www.bfs.admin.ch/bfs/en/home/statistics/population/migration-integration/foreign.html [accessed 2019-07-15] [WebCite Cache ID 784OjJEZI]
46. Schulz P, Hartung U, Riva S. Causes, coping, and culture: a comparative survey study on representation of back pain in three Swiss language regions. PLoS One 2013;8(11):e78029 [FREE Full text] [doi: 10.1371/journal.pone.0078029] [Medline: 24223756]
47. Caiata-Zufferey M, Abraham A, Sommerhalder K, Schulz P. Online health information seeking in the context of the medical consultation in Switzerland. Qual Health Res 2010 Aug;20(8):1050-1061. [doi: 10.1177/1049732310368404] [Medline: 20442347]
48. Carvel J. The Guardian. Patients to rate and review their GPs on NHS website URL: http://www.theguardian.com/society/2008/dec/30/doctors-rating-website-nhs [accessed 2019-07-15] [WebCite Cache ID 6ogmjtvkj]
49. Ärztlicher Kreisverband Ebersberg Körperschaft des öffentlichen Rechts. BÄK Newsletter: Arztbewertungsportale: massive Kritik am AOK-Arzt-Navigator. 2009 URL: http://www.aekv-ebersberg.de/aktuelles/105-baek-newsletter-2009-004-vom-25062009 [accessed 2019-07-15] [WebCite Cache ID 6ogmznsHJ]
50. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [Medline: 15224975]

## Abbreviations

**GP:** general practitioner
**PRW:** physician rating website

*Edited by G Eysenbach; submitted 25.02.19; peer-reviewed by S Bidmon, S Patel, PJ Schulz; comments to author 11.04.19; revised version received 02.05.19; accepted 17.05.19; published 26.07.19.*

*Please cite as:*
*McLennan S*
*Quantitative Ratings and Narrative Comments on Swiss Physician Rating Websites: Frequency Analysis*
*J Med Internet Res 2019;21(7):e13816*
*URL: http://www.jmir.org/2019/7/e13816/*
*doi:10.2196/13816*
*PMID:31350838*

Original Paper

# A Web-Based Exercise System (e-CuidateChemo) to Counter the Side Effects of Chemotherapy in Patients With Breast Cancer: Randomized Controlled Trial

Angelica Ariza-Garcia[1,2*], PhD; Mario Lozano-Lozano[1,3,4,5*], MSc; Noelia Galiano-Castillo[1,3,4,5*], PhD; Paula Postigo-Martin[1,3,5], MSc; Manuel Arroyo-Morales[1,3,4,5], MD, PhD; Irene Cantarero-Villanueva[1,3,4,5], PhD

[1]Department of Physical Therapy, University of Granada, Granada, Spain

[2]"San Cecilio" University Hospital, Granada, Spain

[3]Sport and Health University Research Institute, University of Granada, Granada, Spain

[4]Biohealth Research Institute in Granada, Granada, Spain

[5]Cuídate-Support Unit for Oncology Patients, Granada, Spain

[*]these authors contributed equally

**Corresponding Author:**
Manuel Arroyo-Morales, MD, PhD
Department of Physical Therapy
University of Granada
Avda de la Ilustración, 60
Granada,
Spain
Phone: 34 958248765
Email: marroyo@ugr.es

## Abstract

**Background:** Breast cancer patients have to face a high-risk state during chemotherapy, which involves deterioration of their health including extensive physical deterioration. Face-to-face physical exercise programs have presented low adherence rates during medical treatment, and telehealth systems could improve these adherence rates.

**Objective:** This study aimed to evaluate the effectiveness of a Web-based exercise program (e-CuidateChemo) to mitigate the side effects of chemotherapy on the physical being, anthropometric aspects, and body composition.

**Methods:** A total of 68 patients diagnosed with breast cancer, who were undergoing chemotherapy, were enrolled. The patients were categorized into two groups: e-CuidateChemo (n=34) and controls (n=34). The e-CuidateChemo group participated in an adapted 8-week tailored exercise program through a Web-based system. A blinded, trained researcher assessed functional capacity, strength, anthropometric parameters, and body composition. The intervention effects were tested using analysis of covariance and Cohen *d* tests.

**Results:** Functional capacity improved significantly in the e-CuidateChemo group compared to the control group (6-minute walk test: 62.07 [SD 130.09] m versus –26.34 [SD 82.21] m; 6-minute walk test % distance predicted: 10.81% [SD 22.69%] m versus –4.60% [SD 14.58%]; between-group effect: $P$=.015 for both). The intervention group also showed significantly improved secondary outcomes such as between-group effects for abdominal (24.93 [SD 26.83] s vs –18.59 [SD 38.69] s), back (12.45 [SD 10.20] kg vs 1.39 [10.72] kg), and lower body (–2.82 [SD 3.75] s vs 1.26 [SD 2.84] s) strength; all $P$<.001 compared to the control group.

**Conclusions:** This paper showed that a Web-based exercise program was effective in reversing the detriment in functional capacity and strength due to chemotherapy.

**Trial Registration:** ClinicalTrials.gov NCT02350582; https://clinicaltrials.gov/ct2/show/NCT02350582

***(J Med Internet Res 2019;21(7):e14418)*** doi:10.2196/14418

**KEYWORDS**

breast cancer; chemotherapy; physical fitness; randomized control trial; telehealth; e-health; therapeutic exercise

## Introduction

A diagnosis of cancer is followed by physical and emotional exhaustion that reduces the quality of life. These functional impairments seem to be aggravated with surgery and radiotherapy plus chemotherapy [1] are linked to a decrease in the level of physical activity of up to 50% [2]. The reduction in physical activity is not only an important deterioration of patients' physical capacity [3], but also associated with metabolic changes [4], which increase both the recurrence of cancer and the risk of death [5]. During the treatment, a high-risk period occurs wherein patients with breast cancer become especially sensitive; this period involves a deterioration of health, creating a vicious circle that is difficult to break due to the physical and psychological state of the patients.

There has been a growing interest in rehabilitation through physical activity during cancer treatment in the last few years due to the health related-benefits of such rehabilitation. Patients undergoing chemotherapy find it challenging to maintain a physically active lifestyle during their treatment [2], and physical activity programs following the American College of Sports Medicine guidelines [6] are accepted as effective, safe, and well tolerated in patients with breast cancer who are undergoing chemotherapy [7,8]. These programs focus on aerobic, resistance, and stretching exercises with a moderate-high intensity and could successfully address fatigue and quality of life [9] as well as cardiorespiratory fitness, return to work, and body composition [10]. Furthermore, there may be a positive effect of taking part in physical activity programs to optimize chemotherapy completion rates [11]. It is necessary to emphasize the potential clinical implications of this fact, because greater chemotherapy completion rates may improve disease-free and overall survival [12]. Furthermore, exercise could be the key to counter the effects of chemotherapy and radiation during anticancer treatments [13].

Nevertheless, it is not usual for patients with breast cancer to participate in tailored exercise programs during chemotherapy. Several barriers to exercise in these patients, such as time constraints, confusion regarding the safety of returning to exercise, lack of access to standardized breast cancer–specific exercise programs, or cancer- and treatment-related side effects [14], have been identified. The high costs involved in carrying out on-site physical exercise programs is also linked to this situation. A recent study showed that an on-site physical activity program during chemotherapy is not cost-effective for patients with breast cancer, as such an exercise program accounts for 30% of the total costs [15]. Therefore, alternatives to improve these difficulties are urgently needed, with programs that can be adapted according to each patient in terms of intensity and flexibility.

To combat this issue, current technological advances propose a real alternative that has already shown encouraging results. In our recent study, we found that Web-based systems are effective for improving not only the quality of life, pain, muscle strength, and fatigue [16], but also the functional capacity and cognition [17] in survivors of breast cancer. In addition, this program showed a high rate of adherence (93.9%). In fact, new Web-based systems are also effective in producing behavior changes in terms of diet and physical activity [18]. Nevertheless, few studies have addressed this contemporary topic within the chemotherapy field, the majority of which are nonrandomized controlled trials [19-22]. Most previous experiences with telehealth systems for patients undergoing chemotherapy aimed at self-management, patient assessment, coaching, or alerting a clinician [19], but did not seek specific exercise training or the specific intention to avoid worsening of one's condition. In addition, the literature reveals that self-care systems in patients with chemotherapy are ineffective in managing different symptoms such as fatigue, and therefore, more intervention studies are required to evaluate better strategies for support of cancer patients [23]. A single-arm pilot study [20] proposed a telephone-based exercise intervention to improve fitness, psychological, and anthropometric measures. However, the participants did not receive an adequate tailored intervention [21], or audiovisual material was used [22].

This randomized controlled trial (RCT) aimed to determine the effectiveness of an 8-week low-intensity Web-based therapeutic exercise program for improving the functional capacity, strength, anthropometric parameters, and body composition of patients with breast cancer. We hypothesized that the e-CuidateChemo would prevent the loss of functional capacity and strength and negative changes in anthropometric parameters and body composition after the program in patients with breast cancer undergoing chemotherapy.

## Methods

### Study Design and Participants

This was a two-arm, assessor-blinded, parallel, efficacy RCT (ClinicalTrials.gov: NCT02350582) in which 68 patients with breast cancer undergoing chemotherapy were randomized into the e-CuidateChemo group (n=34) or the control group (n=34). The RCT was performed from September 2013 to June 2015 at a physical therapy laboratory at the University of Granada (Spain). Patients were eligible if they met the following inclusion criteria: diagnosis of stage I-IIIA breast cancer , medical clearance to participate, at the beginning of the chemotherapy, basic ability to use a computer or living with someone who could supervise the first steps using the Web, and having internet access. The participants were excluded if they had a chronic disease or an orthopedic issue that would interfere with the ability to participate in a physical activity program and if they had not provided informed consent.

The Research Ethics Committee of the University of Granada (FIS PI-0457-2010) approved this trial. This trial was performed according to the Helsinki Declaration [24] and the Spanish Biomedical Research Law (14/2007). An oncologist from the chemotherapy unit of the Hospital Virgen de las Nieves (Granada) obtained written informed consent from all participants after the first contact as per the recommendations [25] and recruited the patients according to the established criteria.

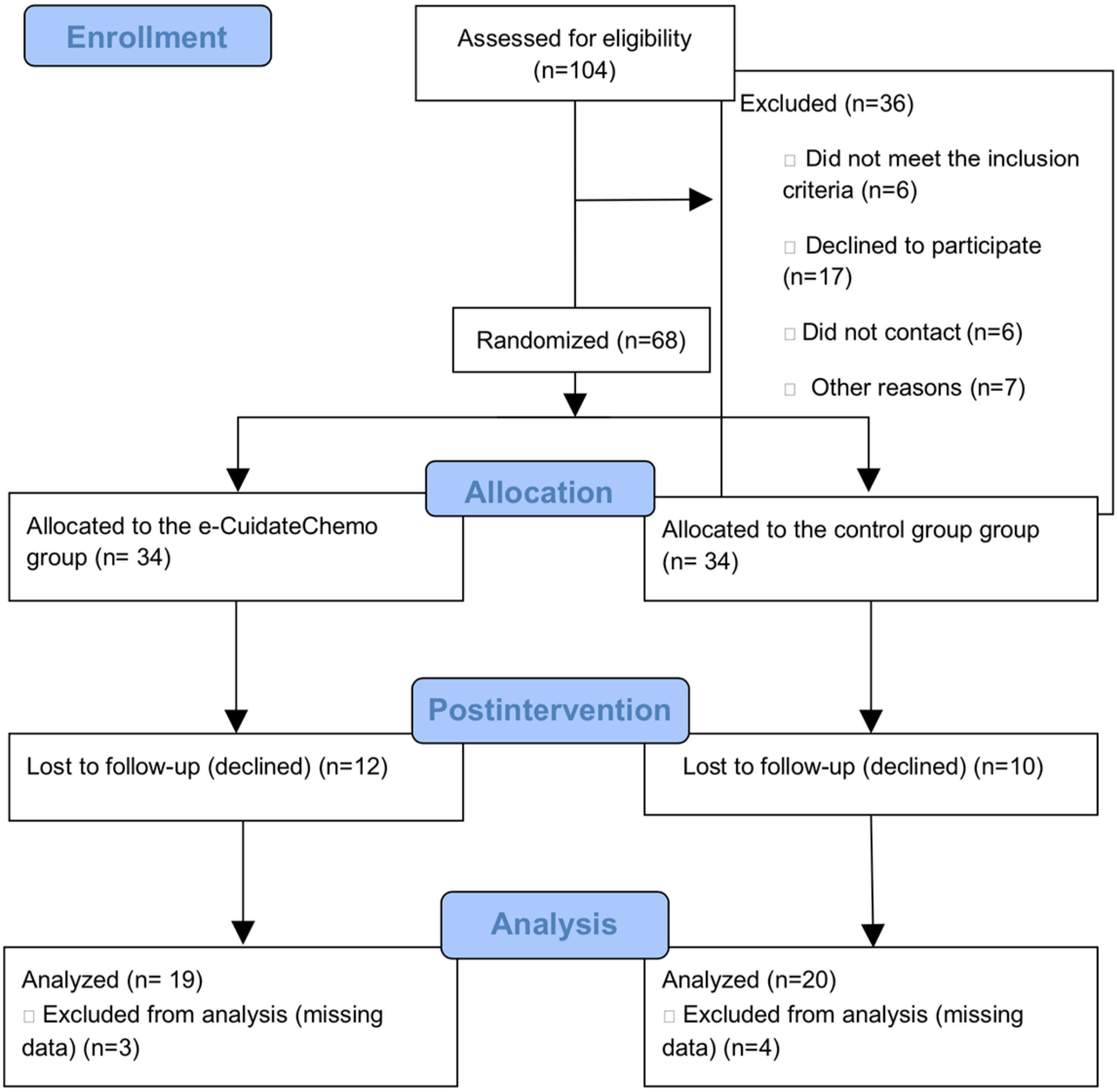
### Randomization and Masking

After completion of the baseline assessment, the eligible patients were randomized into either the e-CuidateChemo group or the control group by using computer-generated numbers (EPIDAT 3.1, Xunta de Galicia Department of Public Department, Coruna, Spain, and Pan American Health Organization, Washington, DC). The researcher in charge of the assessments, with several years of experience with cancer patients, was blinded to the patients' randomization (Figure 1). Thereafter, the sequence was introduced by an external member in sealed opaque envelopes that were opened after the baseline assessment.

### Sample Size Calculation

The sample size and power calculations for this trial were obtained through the overall functional capacity using the 6-minute walk test (6MWT). This was considered the principal outcome of this RCT, especially if we take into account the data reported in a previous study that used a similar online rehabilitation system [17]. The sample size was set at 68 participants (34 per group), providing a 90% power (with a 5% significance) and considering a 30% loss to follow-up due to the specific characteristics of this population. The study recruitment was completed when the predefined sample size was reached.

**Figure 1.** Recruitment and randomization flow diagram process.




### Intervention

The e-CuidateChemo Intervention is a telerehabilitation program that uses an online system [26] adapted to individual requirements. The system is an 8-week program with three sessions per week (on nonconsecutive days). Each session was organized into a warm up, a main, and a cool down part. The aerobic exercise intensity was between 45% and 60% of the maximum heart rate [27] and lasted for 15-30 minutes. There were a total of 5 strength exercises (Table 1 and Textbox 1) of low intensity with functional implementation. The exercises, that is, their volume and intensity, were adapted for each patient according to the baseline assessment. This online system was previously used in breast cancer survivors and showed very high efficiency [16]. As a result, a specific therapeutic exercise program taking into account the special needs of patients with breast cancer who are undergoing chemotherapy has now been developed.

The e-CuidateChemo system also included a communication system between patients and research staff through an internal service. The research team controlled whether the participants received any additional care apart from our intervention. Moreover, weekly contacts were made to ensure correct performance of the intervention and to adapt the intervention to the participants' chemotherapy cycles.

The control group received the usual care with some written basic recommendations for physical exercise, following the general recommendations of the American College of Sports Medicine [6]. The research team controlled the changes in the level of physical activity of these participants through the International Physical Activity Questionnaire. Once the intervention was complete, the control participants were offered participation in the same Web-based exercise program as the intervention group for ethical reasons. No data were recorded from these patients.

**Table 1.** Resistance exercises through the e-CuidateChemo telehealth program.

| Material and week | Time (min) | Volume | Intensity criteria (Borg rating) |
|---|---|---|---|
| **Self-supporting** | | | 10-13[a] |
| 1 | 15 | Training: 1×10-12 repetitions | |
| 2 | 15 | 1×12 repetitions | |
| 3 | 20 | (2×8 repetitions) 30 seconds | |
| 4 | 20 | (2×10 repetitions) 30 seconds | |
| **Elastic bands** | | | 10-13[a] |
| 5 | 20 | Training: 1×12 repetitions | |
| 6 | 25 | (2×12 repetitions) 30 seconds | |
| 7 | 25 | (3×8 repetitions) 30 seconds | |
| 8 | 20 | (2×12 repetitions) 30 seconds | |

[a]Two days with more than a rating of 13 on the Borg scale after training represents a decrease in the intensity of the training program.

**Textbox 1.** Resistance exercises.

Push up: in standing (wall) or prone position (lying down) with/without support

Squat: lifting arms at 90°

Rowing: in semisquat position

Lunge: front and side

Circular movement of the legs in supine position

## Outcomes

All outcomes were assessed at baseline and after the 8-week program. A clinical and sociodemographic questionnaire was used for the assessments.

### *Principal Variable: Functional Capacity*

The 6MWT is a useful measure of functional capacity [28] (H-P-COSMOS for graphics, Germany) [29]. Prior to the test, all patients were familiarized with the treadmill protocol through training performed 2 hours before starting the test and with the period of rest. The participants were instructed to walk as far as possible for 6 minutes [28]. The 6MWT is an objective and reliable test (with an intraclass correlation coefficient [ICC]=0.88) [29]. The results of the 6MWT expressed as percent predicted values were calculated using the reference equation described previously by Enright and colleagues [30]. Finally, participants were distributed for secondary analyses as per their physical activity capacity (normal vs impaired) by using the 75% cutoff, according to Enright and colleagues [30].

### *Outcomes Variables*

#### Abdominal Strength

The abdominal endurance was measured with the patient lying on their back and their knees bent. The patients were instructed to keep the following position as long as possible: arms lifted with the palms guided to the level of the knees, avoiding the lower angle of the scapula from rising from the surface. The research team encouraged the patients and registered the number of seconds they held the position (max of 90 seconds). This is a reliable test with an ICC of 0.97 [31].

#### Lower-Body Strength

In the multiple sit-to-stand test, participants were asked to sit down and stand up from a chair 10 times as fast as possible. The research staff recorded the length of completion of the test in seconds, which had a good reliability, with an ICC of 0.80 [32].

### Lumbar Strength

The lumbar resistance was evaluated using an analog dynamometer (TKK 5002 Back-A, Takey, Tokyo, Japan). The participant was to assume a standing position and maintain a position of 30 degrees. The test was repeated three times, with a 1-minute delay between measurements. Finally, the average of the three measurements was recorded. This test has demonstrated a high reliability, with an ICC of 0.81-0.85 [33].

### Handgrip Strength

A digital dynamometer (TKK 5101 Grip-D, Takey, Tokyo, Japan) with an adjustable grip was used to measure the upper-body muscular strength, registering the average score for each hand (the test was repeated three times with 1-minute delay between measures). This test is valid and reliable [34].

### Anthropometric and Body Composition Outcomes

The waist and hip circumferences were measured using a plastic tape measure. To assess the waist circumference, the plastic tape was placed midway between the lower rib margin and the top of the iliac crest. To measure the hip circumference, the plastic tape was placed at the level of the greater trochanter. These measures have demonstrated a high reliability, with ICCs of 0.89 and 0.81 for the waist and hip circumferences, respectively [35].

We used bioelectrical impedance (InBody 720, Biospace, Gateshead, UK) to measure the body composition. The instrument has a high reliability (ICC=0.98) [36].

## Statistical Analysis

A descriptive analysis was performed, and the mean, 95% confidence interval, and SDs were calculated for each group. To check the differences between groups at the baseline, we used the Student *t* test and Chi-square test. We also used the Chi-square test to calculate the changes in physical activity capacity after the intervention. Normal distribution of the variables was proved with the Shapiro-Wilk test.

Analysis was conducted according to the intention-to-treat principle (with the worst value carried forward in patients who had missing data). The intervention effects on study variables were tested using repeated measure ANCOVA. The time since diagnosis, age, stage of breast cancer, type of surgery, and menopausal status were used as covariates. Regarding the level of significance, interaction effects were reported (5% level of significance). If the analysis revealed a significant interaction, we performed pairwise comparisons with the Bonferroni adjustment to determine if there were differences in the scores between groups. Moreover, the effect size was calculated using Cohen *d* values. The Statistical Program for Social Sciences (version 22.0; IBM, SPSS Statistic for Windows, Armonk, NY) was used for statistical analyses.

# Results

## Sociodemographic and Clinical Data

In summary, 68 patients met the inclusion criteria and were randomized into either the e-CuidateChemo group (n=34; mean age 48.82 [SD 7.68]) or the control group (n=34; mean age 47.32 [SD 9.92]). Figure 1 shows the flow chart of patient distribution and the number and reasons for dropouts. There were 12 dropouts (35.29%) in the e-CuidateChemo group and 10 dropouts in the control group (29.4%). Moreover, three participants from the e-CuidateChemo group and four participants from the control group were excluded from the main analysis due to missing data. Adherence rate for the e-CuidateChemo group, calculated as a ratio of the number of exercise sessions performed in relation to the number of sessions prescribed, was 73.33%. The sociodemographic and medical characteristics are shown in Table 2. None of the participants reported receiving any additional support care in addition to the study program.

**Table 2.** Demographic, clinical, and medical characteristics of the e-CuidateChemo and control groups.

| Characteristic | e-CuidateChemo group (n=34) | Control group (n=34) | *P* value[a] |
|---|---|---|---|
| Age (years), mean (SD) | 48.82 (7.68) | 47.32 (9.92) | .59 |
| **Marital status, n (%)** | | | .51 |
| Single | 4 (18.2) | 2 (9.1) | |
| Married | 16 (72.7) | 16 (72.7) | |
| Divorced/widowed | 2 (9.1) | 4 (18.2) | |
| **Educational level, n (%)** | | | .81 |
| Basic | 9 (40.9) | 8 (36.4) | |
| Medium | 6 (27.3) | 8 (36.4) | |
| High | 7 (31.8) | 6 (27.3) | |
| **Employment status, n (%)** | | | .49 |
| Housewife | 3 (13.6) | 2 (9.1) | |
| Employed | 6 (27.3) | 4 (18.2) | |
| Medical leave/unemployed (by illness) | 12 (54.5) | 12 (54.5) | |
| Unemployed/retired | 1 (4.5) | 4 (18.2) | |
| **Tumor stage, n (%)** | | | .79 |
| I | 7 (31.8) | 5 (22.7) | |
| II | 10 (45.5) | 11 (50.0) | |
| IIIA | 5 (22.7) | 6 (27.3) | |
| **Chemotherapy cycles, n (%)** | | | .44 |
| 4 | 5 (22.7) | 2 (9.1) | |
| 5-7 | 5 (22.7) | 7 (31.8) | |
| 8 | 12 (54.5) | 13 (59.1) | |
| **Type of surgery, n (%)** | | | .48 |
| None | 10 (45.5) | 8 (36.4) | |
| Lumpectomy | 3 (13.6) | 7 (31.8) | |
| Quadrantectomy | 5 (22.7) | 5 (22.7) | |
| Mastectomy | 4 (18.2) | 2 (9.1) | |
| **Menopausal status, n (%)** | | | .22 |
| Premenopausal | 11 (50.0) | 15 (68.2) | |
| Postmenopausal | 11 (50.0) | 7 (31.8) | |
| **International Physical Activity Questionnaire score, n (%)** | | | .89 |
| Low (<500 MET[b]-min/week) | 6 (28.6) | 7 (31.8) | |
| Moderate (500-4499 MET-min/week) | 11 (52.4) | 12 (54.5) | |
| High (≥4500 MET-min/week) | 4 (19.0) | 3 (13.6) | |

[a]*P* values for intergroup comparisons using Student *t* test or Chi-square test, as appropriate.

[b]MET: metabolic equivalent of task.

### Effects of the e-CuidateChemo Intervention on Functional Capacity

Significant interaction effects were found for both 6MWT ($F_{1,37}$=6.51; *P*=.015) and the percentage of 6MWT prediction ($F_{1,37}$=6.44; *P*=.015). We found a significant difference in favor of the intervention group for the walked distance and the 6MWT predicted percentage (both *P*=.015), with distance and percentage increasing in the intervention group but decreasing in the control group. After the intervention, the effect size values were large for both 6MWT (*d*=0.83; 95% CI –32.23 to 33.91) and 6MWT predicted percentage (*d*=0.83; 95% CI –4.96 to 6.63; Table 3). We found no changes in the results after inclusion of the covariates.

**Table 3.** Within-group and between-group effects for physical outcomes at baseline and after the 8-week intervention. Data are shown as mean (SD) and 95% CI for the mean at the baseline and after the 8-week intervention and as the mean difference and 95% CI for the differences for within- and between-group effects.

| Parameter | e-CuidateChemo group (n=19) | Control group (n=20) | Between-group effects |
|---|---|---|---|
| **Functional capacity** | | | |
| **6-minute walk test (m)** | | | |
| Baseline | 421.38 (176.53); 36.30-506.47 | 480.13 (134.98); 416.96-543.30 | N/A[a] |
| 8-week intervention | 483.46 (149.37); 411.46-555.45 | 453.79 (99.98); 406.99-500.59 | N/A |
| Within-group effect - baseline to 8 weeks | 62.07 (130.09); –0.63 to 124.77 | –26.34 (82.21); –64.81 to 12.13 | –88.41; –158.64 to –18.18[b,c] |
| **6-minute walk test % predicted (m)** | | | |
| Baseline | 74.52 (27.32); 61.35-87.69 | 85.23 (23.29); 74.32-96.13 | N/A |
| 8-week intervention | 85.34 (20.33); 75.54-95.14 | 80.62 (16.33); 72.98-88.27 | N/A |
| Within-group effect - baseline to 8 weeks | 10.81 (22.69); –0.11 to 21.75 | –4.60 (14.58); –11.42 to 2.21 | –15.42; –27.73 to –3.11[b,c] |
| **Strength** | | | |
| **Abdominal strength (s)** | | | |
| Baseline | 29.01 (27.29); 15.85-42.16) | 48.60 (46.07); 27.04-70.16 | N/A |
| 8-week intervention | 53.94 (39.03); 35.13-72.76 | 30.01 (17.90); 21.63-38.39 | N/A |
| Within-group effect - baseline to 8 weeks | 24.93 (26.83); 12.00-37.87 | –18.59 (38.69); –36.69 to –0.48 | –43.74; –64.88 to –22.60[c,d] |
| **Lower-body strength (s)** | | | |
| Baseline | 24.30 (4.53); 22.11-26.49 | 23.23 (3.54); 21.57-24.89 | N/A |
| 8-week intervention | 21.47 (3.58); 19.74-23.20 | 24.50 (4.32); 22.48-26.52 | N/A |
| Within-group effect - baseline to 8 weeks | –2.82 (3.75); –4.63 to –1.01 | 1.26 (2.84); –0.06 to 2.60 | 4.11; 2.01-6.21[c,d] |
| **Back strength (kg)** | | | |
| Baseline | 41.05 (15.06); 33.79-48.31 | 39.27 (15.14); 32.18-46.36 | N/A |
| 8-week intervention | 53.50 (16.01); 45.78-61.22 | 40.66 (13.88); 34.16-47.16 | N/A |
| Within-group effect - baseline to 8 weeks | 12.45 (10.20); 7.53-17.37 | 1.39 (10.72); –3.62 to 6.41 | –10.59; –17.27 to –3.90[b,c] |
| **Handgrip strength - affected side (kg)** | | | |
| Baseline | 23.41 (6.62); 20.21-26.60 | 23.76 (3.77); 22.00-25.53 | N/A |
| 8-week intervention | 25.45 (5.94); 22.58-28.32 | 25.08 (4.46); 22.99-27.16 | N/A |
| Within-group effect - baseline to 8 weeks | 2.04 (2.75); 0.71-3.36 | 1.31 (3.70); –0.42 to 3.04 | –0.79; –2.86 to 1.27 |
| **Handgrip strength - nonaffected side (kg)** | | | |
| Baseline | 24.03 (5.01); 21.61-26.44 | 24.72 (4.42); 22.65-26.79 | N/A |
| 8-week intervention | 24.74 (5.00); 22.32-27.15 | 24.70 (4.33); 22.67-26.73 | N/A |
| Within-group effect - baseline to 8 weeks | 0.71 (1.82); –0.16 to 1.59 | –0.01 (2.19); –1.04 to 1.01 | –0.78; –2.07 to 0.49 |

[a]N/A: not applicable.

[b]*P*<.05 (significant between-group effect).

[c]Large effect size: Cohen *d*>0.8.

[d]*P*<.001 (significant between-group effect).

Table 4 shows the differences between patients with a normal physical activity capacity and those with an impaired physical activity capacity after the intervention. There was an increase in the number of participants recovering the normal exercise capacity in the e-CuidateChemo group (45.5% to 78.9%, pre-postintervention) compared to the decrease in the control group (73.3% to 65%). Statistical analysis revealed significant changes between the two groups (*P*=.02).

**Table 4.** Baseline, postintervention, and change in physical exercise capacity. Data are shown as frequencies (percentages) for baseline and postintervention and as frequency differences for change (including loss to follow-up).

| Time point | e-CuidateChemo group | | Control group | |
|---|---|---|---|---|
| | Normal physical exercise capacity | Impaired physical exercise capacity | Normal physical exercise capacity | Impaired physical exercise capacity |
| Baseline, n (%) | 10 (45.5) | 12 (54.5) | 17 (77.3) | 5 (22.7) |
| Postintervention, n (%) | 15 (78.9) | 4 (21.1) | 13 (65.0) | 7 (35.0) |
| Change[a], n | 5 | –8 | –5 | 2 |

[a]Impaired to normal physical exercise capacity.

### Effects of the e-CuidateChemo Intervention on Muscle Strength

The ANCOVA revealed significant interaction effects for abdominal strength ($F_{1,38}$=17.55; $P$<.001) and back strength ($F_{1,38}$=10.28; $P$=.003). The e-CuidateChemo group showed an increase in abdominal and back strength (both $P$<.001) after the intervention compared with the control group, which led to a decrease in abdominal strength and back strength at the baseline (Table 3). We also obtained a significant interaction effect for lower-body strength ($F_{1,38}$=15.74; $P$<.001). In this case, the e-CuidateChemo group showed an improvement in their lower-body strength after the intervention ($P$<.001), while the control group showed similar results at the baseline (Table 3). The intergroup effect size was large for all variables, namely, abdominal strength ($d$=1.33; 95% CI –8.88 to 11.55), lower-body strength ($d$=–1.26; 95% CI –2.27 to –0.251), and back strength ($d$=1.08; 95% CI –2.11 to 4.28). The inclusion of the covariates did not change the results of any of the variables. Regarding the other strength-related measures, handgrip strength for the affected and nonaffected sides did not reveal significant interaction effects ($F_{1,38}$=0.60; $P$=.44 and $F_{1,38}$=1.54; $P$=.22, respectively).

### Effects of the e-CuidateChemo Intervention on Anthropometric Parameters and Body Composition

The repeated-measure ANCOVA analyses did not show any significant interaction effects for any of the variables, namely, waist and hip circumferences, weight, body fat, lean mass, and body mass index (Table 5).

**Table 5.** Within-group and between-group effects for anthropometric and body composition variables at the baseline and after 8-week intervention. Data are shown as mean (SD) and 95% CI for the mean at baseline and 8-week intervention and as mean differences and 95% CI for the differences for within- and between-group effects.

| Parameter | e-CuidateChemo group (n=19) | Control group (n=20) | Between-groups effects |
|---|---|---|---|
| **Waist circumference (cm)** | | | |
| Baseline | 85.20 (11.66); 79.81-90.74 | 86.10 (8.70); 82.02-90.17 | N/A[a] |
| 8-week intervention | 86.01 (11.07); 80.83-91.19 | 86.28 (10.88); 81.19-91.37 | N/A |
| Within-group effect - baseline to 8 weeks | 0.73 (3.41); –0.86 to 2.33 | 0.18 (3.47); –1.43 to 1.80 | –0.55; –2.75 to 1.65 |
| **Hip circumference (cm)** | | | |
| Baseline | 101.58 (9.59); 97.08-106.07 | 103.82 (8.74); 99.72-107.91 | N/A |
| 8-week intervention | 102.72 (8.60); 98.69-106.75 | 103.50 (9.59); 99.01-107.99 | N/A |
| Within-group effect - baseline to 8 weeks | 1.14 (3.20); –0.35 to 2.64 | –0.31 (1.88); –1.19 to 0.56 | –1.46; –3.14 to 0.22 |
| **Weight (kg)** | | | |
| Baseline | 66.46 (12.29); 60.70-72.21 | 67.82 (10.47); 62.91 to 72.72 | N/A |
| 8-week intervention | 67.35 (11.27); 62.07-72.62 | 68.20 (11.71); 62.72 to 73.68 | N/A |
| Within-group effect - baseline to 8 weeks | 0.89 (2.89); –0.46 to 2.24 | 0.38 (2.57); –0.82 to 1.59 | –0.50; –2.26 to 1.25 |
| **Body fat (%)** | | | |
| Baseline | 32.90 (9.60); 28.41-37.39 | 35.01 (7.32); 31.57-38.43 | N/A |
| 8-week intervention | 33.41 (9.01); 29.19-37.63 | 33.13 (6.69); 29.99-36.26 | N/A |
| Within-group effect - baseline to 8 weeks | 0.51 (2.17); –0.50 to 1.52 | –1.87 (7.04); –5.17 to 1.42 | –2.38; –5.72 to 0.95 |
| **Lean mass (kg)** | | | |
| Baseline | 23.84 (2.61); 22.61-25.06 | 23.76 (3.15); 22.28-25.23 | N/A |
| 8-week intervention | 23.93 (2.75); 22.64-25.22 | 23.93 (3.32); 22.37-25.48 | N/A |
| Within-group effect - baseline to 8 weeks | 0.09 (1.05); –0.39 to 0.58 | 0.17 (1.05); –0.32 to 0.66 | 0.07; –0.59 to 0.74 |
| **Body mass index (kg/m$^2$)** | | | |
| Baseline | 26.31 (4.97); 23.98-28.63 | 26.79 (3.79); 25.01-28.56 | N/A |
| 8-week intervention | 26.64 (4.58); 24.49-28.78 | 26.89 (4.32); 24.87-28.91 | N/A |
| Within-group effect - baseline to 8 weeks | 0.33 (1.09); –0.18 to 0.84 | 0.10 (1.02); –0.37 to 0.58 | –0.22; –0.90 to 0.45 |

[a]N/A: not applicable.

## Discussion

The results of this RCT show that a Web-based exercise program is effective in reversing the detriment in functional capacity and strength, which reflects a physical deterioration normally experienced by patients with breast cancer who are undergoing chemotherapy. Having an adequate physical condition during chemotherapy improves the health state of the patients [37-40], reduces side effects, allows modulations of the response to chemotherapy [41], and can even decrease the size of tumors [42]. Therefore, the findings of this RCT provide evidence about an adequate support for patients with breast cancer during chemotherapy.

This therapeutic program involved an improvement, with significant differences in the walked distance of the 6MWT in the e-CuidateChemo group as compared to the control group. The e-CuidateChemo group had a large effect size, showing its effectiveness despite the low intensities and volumes of aerobic exercises (up to a maximum of 30 min at 60% of the maximum heart rate). This increase experienced by the e-CuidateChemo group in the walked distance is above the smaller change considered clinically relevant in cancer patients (43.1 m) [43,44].

In our previous randomized trial with breast cancer survivors [17], a moderate-intensity Web-based exercise program showed an improvement of 104.84 m after 8 weeks. However, the program involved more specific training tailored to improve cardiorespiratory fitness (the American College of Sports Medicine) and the participants had finished a medical treatment. Other previous randomized studies on face-to-face exercise programs also found improvements with aerobic fitness, but all of them involved more specific and intense aerobic exercise programs [45]. The current evidence seems to indicate a higher gain of physical fitness with moderate- or high-intensity exercise programs [46]. A Web-based support system could be useful for encouraging patients to avoid the barriers of exercise [47].

It is also important to emphasize the presence of a significant difference between groups in terms of the 6MWT percentage predicted, which may also be used to identify the decrease in the exercise capacity [30]. We found an increase of 10.8% in the e-CuidateChemo group (with an average change between 74.5% and 85.3%) and a decrease in the control group (change from 85.2% to 80.6%). Within the e-CuidateChemo group, 33.4% of patients with breast cancer reached a normal exercise capacity. In contrast, 12.3% of participants in the control group reduced their physical exercise capacity after 8 weeks. These results were slightly lower than those reported in more specific previous studies [17,45]; in contrast, the results of van Waart and colleagues [11] confirm that low-level physical exercise can help minimize the decline in cardiorespiratory fitness. This fact is very important. Evidence shows that the improvement of cardiorespiratory fitness could help offset the medical treatment–related side effects such as heart damage [48,49]. In addition, it could modulate the response to chemotherapy [41]. Therefore, these data justify the need to integrate such an intervention in the care routine of patients with breast cancer during their treatments.

The Web-based exercise program also achieved an improvement in almost all estimations of muscle strength. The results showed significant differences in abdominal, lower-body, and back strength, with a difference of 43.74 s, 4.11 s, and 10.59 kg, respectively, between both groups. We also found an improvement of 85.9% (abdominal strength), 11.6% (lower-body strength), and 29.74% (back strength) in the e-CuidateChemo group compared with the control group (deterioration of 38.2% and 5.4% and an improvement of only 3.53%, respectively). Thus, the e-CuidateChemo program was successful in limiting the loss of strength of the musculature, which is essential for the development of daily activities, given its relation to the ability to move. These results are of vital importance, since chemotherapy induces wasting, weakness, and muscle fatigue [50], which could be reflected in the loss of strength or a minor improvement seen in the control group. Adams and collaborators [51] showed a reversion of sarcopenia with a moderate-resistance training program (between 60% and 70% of the maximum repetition) and a clinical improvement in the quality of life in patients with breast cancer during chemotherapy. Surprisingly, our program only has five basic resistance exercises of low intensity (ratings of 10-13 on the Borg scale), and these are sufficient to avoid the loss of strength and even improve it.

We did not find significant differences between groups in our handgrip strength or lumbar strength results. The constant use of these muscle in daily activities may have maintained this muscle in both groups. Furthermore, few exercises of these area (Table 1 and Textbox 1) were included in our training routine.

The analysis of our results related to anthropometric parameters and body composition showed a maintenance of these values, following the same trend in the control group. Kim and collaborators found a significant reduction of weight, body mass index, and body fat with a moderate-to-high intensity program based on walking during 12 weeks (5 consecutive days) [52]. However, our Web-based program was not tailored to improve these variables. Avoiding poor values related to these variable could be addressed in future studies, given their influence in lower physical conditions [53], chemotherapy toxicities [10,54], recurrence, or mortality [55,56] in patients with breast cancer.

The main strength of this work is that this is the first study, to our knowledge, that has tested the effectiveness of a low-intensity exercise program based on a Web-based system, to improve physical fitness, anthropometric parameters, and body composition in patients with breast cancer undergoing chemotherapy. Therefore, this study contributes to the current knowledge in this field. The e- CuidateChemo system could be an optimal alternative to support patients with breast cancer, preventing some of the barriers related to their participation in physical exercise programs and saving costs as compared to the high cost of face-to-face programs [15]. Nevertheless, some limitations should be noted. The use of a treadmill to develop the 6MWT is questionable due to a possible overestimation, despite the improvement in terms of percentages predicted in the walked distance. These results should be considered with prudence. Finally, a more extended study throughout the chemotherapy treatment could have produced different results. Thus, further studies are needed to improve the knowledge in this field and to examine whether the observed benefits continue after a long follow-up period.

In conclusion, this low-intensity Web-based exercise program is effective in reversing the detriment of the functional capacity and strength in patients with breast cancer undergoing chemotherapy. The e-CuidateChemo system could be an excellent option to limit the physical deterioration of patients with breast cancer undergoing chemotherapy, because it could prevent the known barriers to practice of physical exercise during chemotherapy [8].

## Acknowledgments

The study was funded by a research project grant from the Andalusian Health Service, Junta de Andalucia, call for subsidies for the financing of biomedical research and health sciences in Andalusia (SAS-0457-2010) and by the Spanish Ministry of Education (FPU14/01069 and FPU17/00939). This study was possible due to the additional funding from the University of Granada, Plan Propio de Investigación 2016, Excellence actions: Units of Excellence; Unit of Excellence on Exercise and Health (UCEES). We are also grateful to Ms Carmen Sainz-Quinn for her assistance with the English language.

## Authors' Contributions

MM conceived the study and designed the e-CuidateChemo program. NC, IV, and AG contributed to the protocol enterprise and designed the diary sessions of the e-CuidateChemo intervention. AG collected all the data and carried out the intervention sessions. NC, MM, and ML planned and carried out the statistical analyses of data and interpreted the results. NC, PM, IV, and ML revised the scientific literature. IV, AG, and MM facilitated communication between the hospital centers and university laboratories. NC, IV, PM, and ML drafted the manuscript. MM advised on the medical aspect of the protocol. NC and AG participated in the enrolment of the patients to the study. All authors read and approved the final manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 542KB - jmir_v21i7e14418_app1.pdf ]

## References


1. Moros M, Ruidiaz M, Caballero A, Serrano E, Martínez V, Tres A. [Effects of an exercise training program on the quality of life of women with breast cancer on chemotherapy]. Rev Med Chil 2010 Jun;138(6):715-722. [Medline: 20919481]
2. Irwin M, Crumley D, McTiernan A, Bernstein L, Baumgartner R, Gilliland F, et al. Physical activity levels before and after a diagnosis of breast carcinoma: the Health, Eating, Activity, and Lifestyle (HEAL) study. Cancer 2003 Apr 01;97(7):1746-1757 [FREE Full text] [doi: 10.1002/cncr.11227] [Medline: 12655532]
3. Ariza-García A, Galiano-Castillo N, Cantarero-Villanueva I, Fernández-Lao C, Díaz-Rodríguez L, Arroyo-Morales M. Influence of physical inactivity in psychophysiological state of breast cancer survivors. Eur J Cancer Care (Engl) 2013 Nov;22(6):738-745. [doi: 10.1111/ecc.12101] [Medline: 23889104]
4. Bhute V, Ma Y, Bao X, Palecek S. The Poly (ADP-Ribose) Polymerase Inhibitor Veliparib and Radiation Cause Significant Cell Line Dependent Metabolic Changes in Breast Cancer Cells. Sci Rep 2016 Dec 04;6:36061 [FREE Full text] [doi: 10.1038/srep36061] [Medline: 27811964]
5. Blanchard C, Courneya K, Stein K, American Cancer Society's SCS-II. Cancer survivors' adherence to lifestyle behavior recommendations and associations with health-related quality of life: results from the American Cancer Society's SCS-II. J Clin Oncol 2008 May 01;26(13):2198-2204. [doi: 10.1200/JCO.2007.14.6217] [Medline: 18445845]
6. Schmitz K, Courneya K, Matthews C, Demark-Wahnefried W, Galvão DA, Pinto B, American College of Sports Medicine. American College of Sports Medicine roundtable on exercise guidelines for cancer survivors. Med Sci Sports Exerc 2010 Jul;42(7):1409-1426. [doi: 10.1249/MSS.0b013e3181e0c112] [Medline: 20559064]
7. Courneya K, McKenzie D, Gelmon K, Mackey JR, Reid R, Yasui Y, et al. A multicenter randomized trial of the effects of exercise dose and type on psychosocial distress in breast cancer patients undergoing chemotherapy. Cancer Epidemiol Biomarkers Prev 2014 May;23(5):857-864 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-1163] [Medline: 24599578]
8. Courneya K, Segal R, Mackey J, Gelmon K, Friedenreich C, Yasui Y, et al. Effects of exercise dose and type on sleep quality in breast cancer patients receiving chemotherapy: a multicenter randomized trial. Breast Cancer Res Treat 2014 Apr;144(2):361-369. [doi: 10.1007/s10549-014-2883-0] [Medline: 24554388]
9. Schmidt M, Wiskemann J, Armbrust P, Schneeweiss A, Ulrich C, Steindorf K. Effects of resistance exercise on fatigue and quality of life in breast cancer patients undergoing adjuvant chemotherapy: A randomized controlled trial. Int J Cancer 2015 Jul 15;137(2):471-480 [FREE Full text] [doi: 10.1002/ijc.29383] [Medline: 25484317]
10. Shachar S, Deal A, Weinberg M, Williams G, Nyrop K, Popuri K, et al. Body Composition as a Predictor of Toxicity in Patients Receiving Anthracycline and Taxane-Based Chemotherapy for Early-Stage Breast Cancer. Clin Cancer Res 2017 Jul 15;23(14):3537-3543 [FREE Full text] [doi: 10.1158/1078-0432.CCR-16-2266] [Medline: 28143874]
11. van Waart H, Stuiver M, van Harten WH, Geleijn E, Kieffer J, Buffart L, et al. Effect of Low-Intensity Physical Activity and Moderate- to High-Intensity Physical Exercise During Adjuvant Chemotherapy on Physical Fitness, Fatigue, and Chemotherapy Completion Rates: Results of the PACES Randomized Clinical Trial. J Clin Oncol 2015 Jun 10;33(17):1918-1927. [doi: 10.1200/JCO.2014.59.1081] [Medline: 25918291]
12. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Peto R, Davies C, Godwin J, Gray R, Pan HC, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet 2012 Feb 04;379(9814):432-444 [FREE Full text] [doi: 10.1016/S0140-6736(11)61625-5] [Medline: 22152853]
13. Lemanne D, Cassileth B, Gubili J. The role of physical activity in cancer prevention, treatment, recovery, and survivorship. Oncology (Williston Park) 2013 Jun;27(6):580-585 [FREE Full text] [Medline: 23909073]
14. Schmitz K. Exercise for secondary prevention of breast cancer: moving from evidence to changing clinical practice. Cancer Prev Res (Phila) 2011 Apr;4(4):476-480 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-11-0097] [Medline: 21464028]
15. May A, Bosch M, Velthuis M, van der Wall E, Steins Bisschop CN, Los M, et al. Cost-effectiveness analysis of an 18-week exercise programme for patients with breast and colon cancer undergoing adjuvant chemotherapy: the randomised PACT study. BMJ Open 2017 Dec 06;7(3):e012187 [FREE Full text] [doi: 10.1136/bmjopen-2016-012187] [Medline: 28264824]

16. Galiano-Castillo N, Cantarero-Villanueva I, Fernández-Lao C, Ariza-García A, Díaz-Rodríguez L, Del-Moral-Ávila R, et al. Telehealth system: A randomized controlled trial evaluating the impact of an internet-based exercise intervention on quality of life, pain, muscle strength, and fatigue in breast cancer survivors. Cancer 2016 Jun 22;122(20):3166-3174. [doi: 10.1002/cncr.30172]
17. Galiano-Castillo N, Arroyo-Morales M, Lozano-Lozano M, Fernández-Lao C, Martín-Martín L, Del-Moral-Ávila R, et al. Effect of an Internet-based telehealth system on functional capacity and cognition in breast cancer survivors: a secondary analysis of a randomized controlled trial. Support Care Cancer 2017 Dec;25(11):3551-3559. [doi: 10.1007/s00520-017-3782-9] [Medline: 28639097]
18. Lee MK, Yun YH, Park H, Lee ES, Jung KH, Noh D. A Web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
19. Beck S, Eaton L, Echeverria C, Mooney K. SymptomCare@Home: Developing an Integrated Symptom Monitoring and Management System for Outpatients Receiving Chemotherapy. Comput Inform Nurs 2017 Oct;35(10):520-529 [FREE Full text] [doi: 10.1097/CIN.0000000000000364] [Medline: 28570285]
20. Ligibel J, Partridge A, Giobbie-Hurder A, Campbell N, Shockro L, Salinardi T, et al. Physical and psychological outcomes among women in a telephone-based exercise intervention during adjuvant therapy for early stage breast cancer. J Womens Health (Larchmt) 2010 Aug;19(8):1553-1559. [doi: 10.1089/jwh.2009.1760] [Medline: 20629579]
21. Lustria M, Noar S, Cortese J, Van Stee SK, Glueckauf R, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
22. Tuong W, Larsen E, Armstrong A. Videos to influence: a systematic review of effectiveness of video-based education in modifying health behaviors. J Behav Med 2014 Apr;37(2):218-233. [doi: 10.1007/s10865-012-9480-7] [Medline: 23188480]
23. Richardson A, Ream E. Self-care behaviours initiated by chemotherapy patients in response to fatigue. Int J Nurs Stud 1997 Feb;34(1):35-43. [Medline: 9055119]
24. 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]
25. Park J, Lee J, Oh M, Park H, Chae J, Kim D, et al. The effect of oncologists' exercise recommendations on the level of exercise and quality of life in survivors of breast and colorectal cancer: A randomized controlled trial. Cancer 2015 Aug 15;121(16):2740-2748 [FREE Full text] [doi: 10.1002/cncr.29400] [Medline: 25965782]
26. CUIDATE. URL: http://wpd.ugr.es/~unidadcuidate/wordpress/ [accessed 2019-06-27]
27. Nes BM, Janszky I, Wisløff U, Støylen A, Karlsen T. Age-predicted maximal heart rate in healthy subjects: The HUNT Fitness Study. Scand J Med Sci Sports 2012 Feb 29;23(6):697-704. [doi: 10.1111/j.1600-0838.2012.01445.x]
28. Holland A, Spruit M, Troosters T, Puhan M, Pepin V, Saey D, et al. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. Eur Respir J 2014 Dec;44(6):1428-1446 [FREE Full text] [doi: 10.1183/09031936.00150314] [Medline: 25359355]
29. Laskin J, Bundy S, Marron H, Moore H, Swanson M, Blair M, et al. Using a treadmill for the 6-minute walk test: reliability and validity. J Cardiopulm Rehabil Prev 2007;27(6):407-410. [doi: 10.1097/01.HCR.0000300270.45881.d0] [Medline: 18197077]
30. Enright PL, Sherrill DL. Reference equations for the six-minute walk in healthy adults. Am J Respir Crit Care Med 1998 Nov;158(5 Pt 1):1384-1387. [doi: 10.1164/ajrccm.158.5.9710086] [Medline: 9817683]
31. McGill S, Childs A, Liebenson C. Endurance times for low back stabilization exercises: clinical targets for testing and training from a normal database. Arch Phys Med Rehabil 1999 Aug;80(8):941-944. [Medline: 10453772]
32. Ritchie C, Trost S, Brown W, Armit C. Reliability and validity of physical fitness field tests for adults aged 55 to 70 years. J Sci Med Sport 2005 Mar;8(1):61-70. [Medline: 15887902]
33. Imagama S, Matsuyama Y, Hasegawa Y, Sakai Y, Ito Z, Ishiguro N, et al. Back muscle strength and spinal mobility are predictors of quality of life in middle-aged and elderly males. Eur Spine J 2011 Jun;20(6):954-961 [FREE Full text] [doi: 10.1007/s00586-010-1606-4] [Medline: 21072545]
34. Jakobsen L, Rask I, Kondrup J. Validation of handgrip strength and endurance as a measure of physical function and quality of life in healthy subjects and patients. Nutrition 2010 May;26(5):542-550. [doi: 10.1016/j.nut.2009.06.015] [Medline: 19804953]
35. Sonnenschein E, Kim M, Pasternack B, Toniolo P. Sources of variability in waist and hip measurements in middle-aged women. Am J Epidemiol 1993 Sep 01;138(5):301-309. [doi: 10.1093/oxfordjournals.aje.a116859] [Medline: 8356968]
36. Thomas E, Frost G, Harrington T, Bell J. Validation of 'InBody' Bioelectrical Impedance by Whole Body MRI. Laboratory Report 2001:1-2 [FREE Full text]
37. Cheema B, Gaul C, Lane K, Fiatarone Singh MA. Progressive resistance training in breast cancer: a systematic review of clinical trials. Breast Cancer Res Treat 2008 May;109(1):9-26. [doi: 10.1007/s10549-007-9638-0] [Medline: 17624588]
38. Knols R, Aaronson N, Uebelhart D, Fransen J, Aufdemkampe G. Physical exercise in cancer patients during and after medical treatment: a systematic review of randomized and controlled clinical trials. J Clin Oncol 2005 Jun 01;23(16):3830-3842. [doi: 10.1200/JCO.2005.02.148] [Medline: 15923576]

39. Kuchinski A, Reading M, Lash A. Treatment-related fatigue and exercise in patients with cancer: a systematic review. Medsurg Nurs 2009;18(3):174-180. [Medline: 19591364]
40. Velthuis M, Agasi-Idenburg S, Aufdemkampe G, Wittink H. The effect of physical exercise on cancer-related fatigue during cancer treatment: a meta-analysis of randomised controlled trials. Clin Oncol (R Coll Radiol) 2010 Apr;22(3):208-221. [doi: 10.1016/j.clon.2009.12.005] [Medline: 20110159]
41. Jones L, Fels D, West M, Allen J, Broadwater G, Barry W, et al. Modulation of circulating angiogenic factors and tumor biology by aerobic training in breast cancer patients receiving neoadjuvant chemotherapy. Cancer Prev Res (Phila) 2013 Sep;6(9):925-937 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-12-0416] [Medline: 23842792]
42. Dethlefsen C, Pedersen KS, Hojman P. Every exercise bout matters: linking systemic exercise responses to breast cancer control. Breast Cancer Res Treat 2017 Dec;162(3):399-408. [doi: 10.1007/s10549-017-4129-4] [Medline: 28138894]
43. Granger C, Holland A, Gordon I, Denehy L. Minimal important difference of the 6-minute walk distance in lung cancer. Chron Respir Dis 2015 May;12(2):146-154. [doi: 10.1177/1479972315575715] [Medline: 25749346]
44. Schmidt K, Vogt L, Thiel C, Jäger E, Banzer W. Validity of the six-minute walk test in cancer patients. Int J Sports Med 2013 Jul;34(7):631-636. [doi: 10.1055/s-0032-1323746] [Medline: 23444095]
45. Mutrie N, Campbell A, Whyte F, McConnachie A, Emslie C, Lee L, et al. Benefits of supervised group exercise programme for women being treated for early stage breast cancer: pragmatic randomised controlled trial. BMJ 2007 Mar 10;334(7592):517 [FREE Full text] [doi: 10.1136/bmj.39094.648553.AE] [Medline: 17307761]
46. Toohey K, Pumpa K, Arnolda L, Cooke J, Yip D, Craft P, et al. A pilot study examining the effects of low-volume high-intensity interval training and continuous low to moderate intensity training on quality of life, functional capacity and cardiovascular risk factors in cancer survivors. PeerJ 2016;4:e2613 [FREE Full text] [doi: 10.7717/peerj.2613] [Medline: 27781180]
47. McCue M, Fairman A, Pramuka M. Enhancing quality of life through telerehabilitation. Phys Med Rehabil Clin N Am 2010 Feb;21(1):195-205. [doi: 10.1016/j.pmr.2009.07.005] [Medline: 19951786]
48. Ma Z. Effect of anthracycline combined with aerobic exercise on the treatment of breast cancer. Pak J Pharm Sci 2018 May;31(3(Special)):1125-1129. [Medline: 29735461]
49. Jones L, Eves N, Courneya K, Chiu B, Baracos V, Hanson J, et al. Effects of exercise training on antitumor efficacy of doxorubicin in MDA-MB-231 breast cancer xenografts. Clin Cancer Res 2005 Sep 15;11(18):6695-6698 [FREE Full text] [doi: 10.1158/1078-0432.CCR-05-0844] [Medline: 16166449]
50. Gilliam L, St Clair DK. Chemotherapy-induced weakness and fatigue in skeletal muscle: the role of oxidative stress. Antioxid Redox Signal 2011 Nov 01;15(9):2543-2563 [FREE Full text] [doi: 10.1089/ars.2011.3965] [Medline: 21457105]
51. Adams S, Segal R, McKenzie D, Vallerand J, Morielli A, Mackey J, et al. Impact of resistance and aerobic exercise on sarcopenia and dynapenia in breast cancer patients receiving adjuvant chemotherapy: a multicenter randomized controlled trial. Breast Cancer Res Treat 2016 Dec;158(3):497-507. [doi: 10.1007/s10549-016-3900-2] [Medline: 27394134]
52. Kim J, Shin Y, Suk M. Effect of a 12-week walking exercise program on body composition and immune cell count in patients with breast cancer who are undergoing chemotherapy. J Exerc Nutrition Biochem 2015 Sep;19(3):255-262 [FREE Full text] [doi: 10.5717/jenb.2015.15092812] [Medline: 26525495]
53. Cantarero-Villanueva I, Galiano-Castillo N, Fernández-Lao C, Diaz-Rodríguez L, Fernández-Pérez AM, Sánchez MJ, et al. The influence of body mass index on survival in breast cancer patients. Clin Breast Cancer 2015 Apr;15(2):e117-e123. [doi: 10.1016/j.clbc.2014.11.006] [Medline: 25499695]
54. Chan D, Vieira A, Aune D, Bandera EV, Greenwood DC, McTiernan A, et al. Body mass index and survival in women with breast cancer-systematic literature review and meta-analysis of 82 follow-up studies. Ann Oncol 2014 Oct;25(10):1901-1914 [FREE Full text] [doi: 10.1093/annonc/mdu042] [Medline: 24769692]
55. Majed B, Moreau T, Asselain B, Curie Institute Breast Cancer Group. Overweight, obesity and breast cancer prognosis: optimal body size indicator cut-points. Breast Cancer Res Treat 2009 May;115(1):193-203. [doi: 10.1007/s10549-008-0065-7] [Medline: 18546073]
56. Rock C, Byers T, Colditz G, Demark-Wahnefried W, Ganz P, Wolin K, ExerciseNutrition to Enhance RecoveryGood Health for You (ENERGY) Trial Group. Reducing breast cancer recurrence with weight loss, a vanguard trial: the Exercise and Nutrition to Enhance Recovery and Good Health for You (ENERGY) Trial. Contemp Clin Trials 2013 Mar;34(2):282-295 [FREE Full text] [doi: 10.1016/j.cct.2012.12.003] [Medline: 23266440]


## Abbreviations

**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 17.04.19; peer-reviewed by F Artacho-Cordon, Z Zhou; comments to author 15.06.19; revised version received 18.06.19; accepted 18.06.19; published 24.07.19.*

*Please cite as:*
*Ariza-Garcia A, Lozano-Lozano M, Galiano-Castillo N, Postigo-Martin P, Arroyo-Morales M, Cantarero-Villanueva I*
*A Web-Based Exercise System (e-CuidateChemo) to Counter the Side Effects of Chemotherapy in Patients With Breast Cancer: Randomized Controlled Trial*
*J Med Internet Res 2019;21(7):e14418*
*URL: http://www.jmir.org/2019/7/e14418/*
*doi:10.2196/14418*
*PMID:31342907*

Original Paper

# The Impact of Monetary Incentives on Physician Prosocial Behavior in Online Medical Consulting Platforms: Evidence From China

Dong Jing[1], PhD; Yu Jin[2], PhD; Jianwei Liu[2], PhD

[1]Key Laboratory of Interactive Media Design and Equipment Service Innovation, Harbin Institute of Technology, Harbin, China

[2]School of Management, Harbin Institute of Technology, Harbin, China

**Corresponding Author:**
Jianwei Liu, PhD
School of Management
Harbin Institute of Technology
No. 92 Xidazhi Street
Harbin,
China
Phone: 86 15124572304
Email: jianweiliu@hit.edu.cn

## Abstract

**Background:** In online medical consulting platforms, physicians can get both economic and social returns by offering online medical services, such as answering questions or sharing health care knowledge with patients. Physicians' online prosocial behavior could bring many benefits to the health care industry. Monetary incentives could encourage physicians to engage more in online medical communities. However, little research has studied the impact of monetary incentives on physician prosocial behavior and the heterogeneity of this effect.

**Objective:** This study aims to explore the effects of monetary incentives on physician prosocial behavior and investigate the moderation effects of self-recognition and recognition from others of physician competence.

**Methods:** This study was a fixed-effect specification-regression model based on a difference-in-differences design with robust standard errors clustered at the physician level using monthly panel data. It included 26,543 physicians in 3851 hospitals over 133 months (November 2006-December 2017) from a leading online health care platform in China. We used the pricing strategy of physicians and satisfaction levels to measure their own and patients' degree of recognition, respectively. Physicians' prosocial behavior was measured by free services offered.

**Results:** The introduction of monetary incentives had a positive effect on physician prosocial behavior ($\beta$=1.057, *P*<.01). Higher self-recognition and others' recognition level of physician competence increased this promotion effect ($\gamma$=0.275, *P*<.01 and $\gamma$=0.325, *P*<.01).

**Conclusions:** This study explored the positive effect of the introduction of monetary incentives on physician prosocial behavior. We found this effect was enhanced for physicians with a high level of self-recognition and others' recognition of their competence. We provide evidence of the effect of monetary incentives on physicians' prosocial behaviors in the telemedicine markets and insight for relevant stakeholders into how to design an effective incentive mechanism to improve physicians' prosocial engagements.

***(J Med Internet Res 2019;21(7):e14685)*** doi:10.2196/14685

**KEYWORDS**

online medical consultation; monetary incentives; physicians' prosocial behaviors; self-recognition; others' recognition

## Introduction

Online medical service platforms are growing rapidly. They have been viewed as an important supplement for the offline health care industry through medical resource allocation and physician-patient interaction [1]. Online medical consultation services comprise a novel channel through which physicians can offer more intense interactions to patients at lower costs [2,3]. On online health care platforms, physicians answer medical questions and share health care knowledge with patients in their spare time. Meanwhile, physicians can get both economic and social returns by offering their medical services, such as answering questions and sharing health care knowledge [4]. Therefore, if the patients live in remote areas far from a hospital or they need to go to the hospital, the online health care service can provide a more convenient and easy approach for patients to access health services. Online medical interactions, including both paid and free services, could promote increased respect and trust [5,6]. Physicians can build their personal reputation and receive recognition for providing free online medical consulting services [7]. This type of online prosocial behavior not only benefits physicians in a nonprofit way, such as through social returns, but also patients who express their health concerns online [2]. Therefore, online health care services extend traditional offline health services and satisfies the unfulfilled medical demands that offline health care fails to accomplish.

There are still many challenges in the development of the online health care industry. First, choosing an appropriate physician is critical for online patients [8]. However, because of information asymmetry and lack of professional health care knowledge, it is difficult for patients to ascertain a physician's competency and service quality based on limited information and knowledge [9]. A physician's online prosocial behavior could provide such information to help patients in the telemedicine market. Second, although physician online prosocial behavior is an indispensable resource for the development of telemedicine markets [4], participating in and contributing to the online medical marketplace is burdensome for physicians due to their heavy offline workloads. Thus, both patients and physicians encounter difficulties in participating in the online medical marketplace. Understanding how to enable physicians to make more online prosocial contributions has become a managerial agenda for telemedicine practitioners.

The introduction of monetary incentives may influence physicians' prosocial behaviors through self-determination and image concerns [10]. Monetary incentives are increasingly adopted as a method of improving individual performance in many research domains [11-15]. Patients who are satisfied with a physician's online service can pay a service fee to the physician. This monetary incentive can improve the reliability of service [6]. Moreover, this type of incentive can bring both reputational and monetary rewards for physicians, motivate their online contributions, and enhance their service quality. Therefore, introducing monetary incentives might have a positive effect on physician prosocial behavior in the telemedicine market.

Although physician online prosocial behavior has significance for online patients and society, there has been little research to explore it deeply. First, research on physician online prosocial behavior in telemedicine markets is scant. Previous research has been in a wide range of disciplines, such as economics [16] and marketing [17], but it has neglected the existence of the emerging telemedicine context. Second, although there has been extensive research exploring various factors for prosocial behavior—including situational factors, bystander effects [18], and individual factors such as cognitive capacities [19]—little research has investigated the impact of monetary incentives on physician online prosocial behavior. Furthermore, exploration on the effect of monetary incentives on physician behavior can give us a better understanding on the development of the online health care market. Physicians provide consultation services, knowledge, and information to help patients understand their diseases and obtain treatment, which can promote the development of telemedicine markets. Hence, it is important to investigate the role of monetary incentives on physicians' online prosocial contributions. To fill these research gaps, the main research questions leading this study are can monetary incentives improve physicians' prosocial behaviors in online medical consulting platforms and does this effect differ in the extent of physicians' self-recognition and patients' recognition?

# Methods

## Research Hypotheses

Our study investigated the impact of monetary incentives on physicians' prosocial behaviors based on self-determination and image concerns theories [10].

Self-determination theory details intrinsic motivation and extrinsic motivation [10]. Intrinsic motivation means that one is motivated by one's interest in an activity and inherent satisfaction, and extrinsic motivation refers to one's behavior initiated and maintained by contingencies external to the person, such as tangible rewards and intangible rewards. Image concerns refer to an individual's concerns with the perceptions of others. If individuals desire to be liked and respected by others, they would try to adjust their behaviors to signal good traits [12,16].

Generally, prosocial behavior is defined as a contributors' actions that benefit other people or society [20]. Although evidence indicates that intrinsic motivation and extrinsic motivation have separate effects on prosocial behavior [15], they also interact with each other [21,22], especially when monetary incentives are introduced [11,23-25]. In particular, the motivation crowding-out theory shows that monetary incentives may have a negative effect on prosocial behavior by underlying intrinsic motivation [16]. Moreover, monetary incentives also induce extrinsic motivation and bring image concerns, according to previous studies [12,16,26]. In particular, if contributors receive extrinsic rewards for prosocial behaviors, they are suspected of acting prosocially primarily for financial reward rather than out of intrinsic motivation, such as pure altruism or concern for others' well-being. The presence of extrinsic incentives spoils the presentation of a prosocial image and creates doubts regarding the contributors' good deeds.

When making decisions, physicians would predict the outcomes of choosing different actions and may seek to draw lessons from consequences suffered both by themselves and others [27]. To avoid putting themselves in situations of image concern [28-30] or social pressure [31,32], they will care more about appearing prosocial to themselves. If individuals are looking to keep their good image and social approval, they would choose to engage in prosocial activities and contribute more prosocial behaviors after accepting extrinsic rewards. Usually, image concerns may be more dominant than crowding-out effects in social interaction settings.

In telemedicine markets, patients usually communicate with physicians to obtain medical advice and professional treatment, and physicians contribute free feedback to patients to promote their online presence and image [6,33]. Recently, a new type of service feedback has been applied to the telemedicine platforms—paid feedback. Patients may pay service fees to the physicians to encourage them to engage in online medical feedback. However, concentrating on paid feedback means that physicians' online medical services may be just for monetary reward, thereby spoiling the signal of a prosocial image because free feedback indicates more about physicians' prosocial tendencies. Therefore, after accepting monetary incentives, physicians would contribute more free feedback and increase their prosocial behaviors to display private preferences for others' well-being and to avoid looking selfish and greedy. Thus, we hypothesized that the introduction of monetary incentives will improve physicians' prosocial behaviors (hypothesis 1). However, the introduction of monetary incentives may not be equally important for all physicians and may differ by the extent of basic psychological needs.

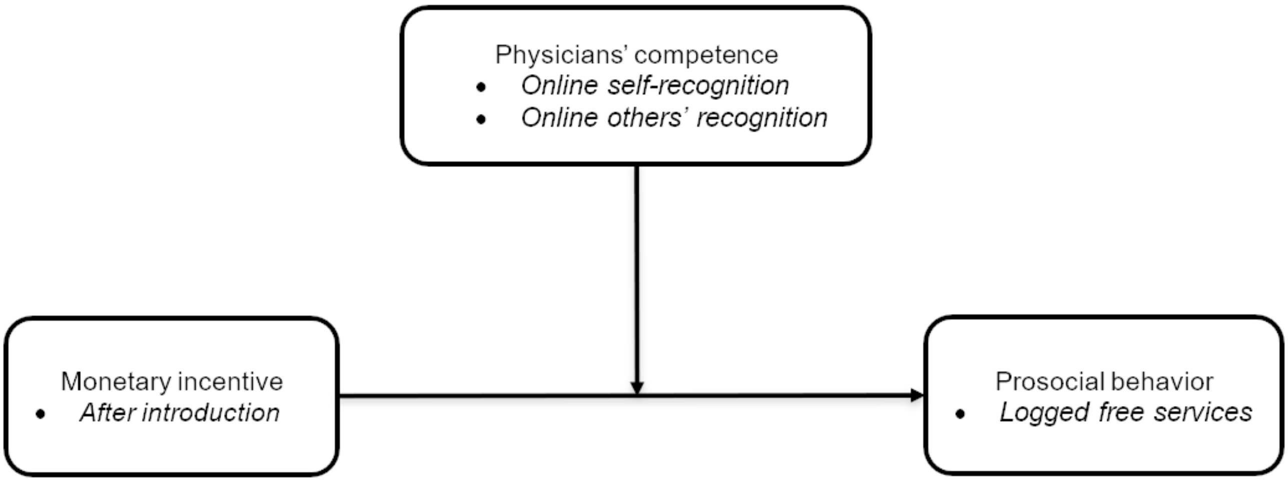
Self-determination theory proposes that human beings have basic psychological needs for autonomy, competence, and relatedness, and that satisfaction of these basic psychological needs provides the nutriments for intrinsic motivation and internalization of extrinsic motivation [10]. Therefore, work climates that support the satisfaction of these needs will promote a person's enjoyment of activities (intrinsic motivation) and the autonomous self-regulation of behaviors (internalization of extrinsic motivation) [34]. Human behaviors can be characterized and determined in terms of the degree of intrinsic motivation and extrinsic motivation [35]. Intrinsic motivation will facilitate good work outcomes, such as effective performance and positive work-related attitudes. Gagné [36] showed that satisfaction of needs will orient people toward paying more attention to others, thus making them more likely to engage in prosocial behaviors.

Toubia and Stephen [30] suggest that competence should encompass the sense of self-worth and social acceptance based on a user's activities on online platforms. In other words, competence is evaluated by oneself and others. Therefore, satisfaction of competence is based on self-recognition and others' recognition of a user's ability and performance. In telemedicine markets, physicians offer online medical consulting services to help patients understand their diseases and get treatment. The competence satisfaction of physicians is determined by their feelings of competence to master online feedback and provide professional treatment, and patient recognition of their past work performance [9]. According to self-determination theory, physicians' prosocial behavior may vary with the level of satisfaction of basic psychological needs (eg, competence satisfaction). Physicians have higher satisfaction in their competence if they obtain higher self-recognition and recognition from others of their competence. Therefore, they are more likely to pay attention to others, thereby contributing free feedback and behaving prosocially. Based on this, we hypothesized the following: the extent of online self-recognition of physicians' competence strengthens the effect of monetary incentives on their prosocial behaviors (hypothesis 2) and the extent of online others' recognition of physicians' competence strengthens the effect of monetary incentives on their prosocial behaviors (hypothesis 3). Our research framework is shown in Figure 1.

## Research Design

The variance in the timing of monetary incentive appearance across physicians provides a unique quasi-experimental opportunity to estimate its influence on physician's online prosocial behavior. With the entry of monetary incentives in a particular month as the treatment, physicians who had at least one entry were the treatment group (ie, physicians with incentive), and those without any entry were the control group (ie, physicians without incentive). We used a difference-in-differences (DID) approach to represent the quasi-experiment [37]. In our DID design, the first difference was between treated physicians and control physicians, and the second difference was between the periods before and after incentive. The double differencing eliminated the potential biases that may come from inherent trends in the prosocial behaviors of physicians.

**Figure 1.** Research framework of the impact of introducing monetary incentives on physician online prosocial behavior.



### Data and Variables

This study collected data from a leading online health care platform called Haodf in China. The functions of the platform include online medical consultation, appointment referral, medical information inquiries, knowledge sharing of medical science, physician recommendations, and so forth. In addition, this platform offers a unique institutional setting to separate free and paid consultations. We identified a paid consultation service as a significant sign of payment as shown in Figure 2. In a free consultation, patients received slow and limited responses from physicians; in a paid consultation, patients communicated immediately with physicians. We selected 26,543 physicians in 3851 hospitals who had provided paid consultation services as the target sample. We developed a crawler to collect the historical data of physicians from November 2006 to December 2017 (133 months) and their attributed data on January 2018 from Haodf. The definitions and statistical descriptions of major variables are shown in Table 1. Additionally, Table 2 presents the correlations of the main variables in the research model, which indicates that there was no significant multicollinearity among the independent variables.

#### *Dependent Variables and Independent Variables*

Physicians' prosocial behaviors were the free consulting services offered by physicians in an online health care community, measured by the logged volume of free answers in a given month as the dependent variable. The introduction of online monetary incentives was the key independent variable of interest in our estimation.

#### *Moderators and Control Variables*

To explore the heterogeneous effects of monetary incentives on physicians' prosocial behavior, we introduced two streams of moderators, including high price and high rated. High price indicated the extent of physicians' self-recognition measured by the pricing strategy of consulting established by physicians. High rated was the extent of others' recognition of a physician measured by the online rating posted by patients. Several control variables were considered to ensure the model robustness; examples include patient votes, letters of thanks, affiliated hospital level, and professional title.

### Research Model

For a physician *i* in month *t*, we modeled the entry effect of monetary incentives as follows:

$$Y_{it} = \beta(\text{monetary incentive})_{it} + \gamma(\text{monetary incentive})_{it} \times Z_{it} + \mu_i + v_t + \varepsilon_{it}$$

where *Y* is the logarithm of monthly free consulting services (prosocial behavior). The monetary incentive dummy variable indicates whether physician *i* has experienced at least one monetary incentive. *Z* is a vector of moderators, including high price and high rated. We account for the unobserved heterogeneity across physician and temporal trends that may be correlated with both monetary incentives and the prosocial behaviors of physicians; *μ* represents the physician-level fixed effects to account for time-invariant characteristics of physicians, *v* is a vector consisting of both month trends and year-month fixed effects to control for temporal trends or shocks that apply to the online medical market, and ε is an idiosyncratic error term. The treatment effect and moderating effects are identified by the coefficient β and the vector of coefficients γ, respectively. We clustered robust standard errors at the physician level to account for the potential correlation in the standard errors within physicians [38,39].

**Figure 2.** Description of the paid consultation services.

**Table 1.** Definitions and summary statistics of variables (N=777,110).

| Variable | Definition | Mean (SD) | Range |
|---|---|---|---|
| **Dependent variable** | | | |
| Prosocial behavior | Logged number of free services offered by a physician in a given month | 1.600 (1.958) | 0-10.788 |
| **Independent variable** | | | |
| Monetary incentive | Dummy variable indicating whether a physician started receiving actual monetary income in a given month | 0.307 (0.461) | 0-1.000 |
| **Moderators** | | | |
| High price[a] | Dummy variable indicating whether the consulting price is high for a physician in online medical consulting platform | 0.279 (0.449) | 0-1 |
| High rated[a] | Dummy variable indicating whether a physician is high rated by the users in online medical consulting platform | 0.451 (0.498) | 0-1 |
| **Control variables** | | | |
| Patient votes | Logged number of votes showing praise given by patients to a physician | 2.815 (1.479) | 0-7.720 |
| Letters of thanks | Logged number of letters of thanks that a physician received in a given month | 0.162 (0.471) | 0-5.273 |
| High hospital level[a] | Dummy variable indicating whether a hospital is designated by the Chinese government as a “third-level grade-A” level | 0.765 (0.424) | 0-1 |
| Professional title | Official clinic title certified by the national agency with uniform standards; Four stages exist for clinic titles: archiater (4), associate archiater (3), chief physician (2), resident physician (1), and none (0). | 3.073 (0.886) | 1-4 |

[a]High price, high rated, and high hospital level are split by their mean values.

**Table 2.** Statistical analysis of pairwise correlation of variables.

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Prosocial behavior | — | | | | | | | |
| 2. Monetary incentive | .461[a] | — | | | | | | |
| 3. Votes | .407[a] | .215[a] | — | | | | | |
| 4. Letters | .497[a] | .34[a] | .353[a] | — | | | | |
| 5. High price | .195[a] | .106[a] | .446[a] | .175[a] | — | | | |
| 6. Title | .035[a] | −.048[a] | .371[a] | .022[a] | .229[a] | — | | |
| 7. High rated | .214[a] | .127[a] | .603[a] | .222[a] | .350[a] | .246[a] | — | |
| 8. Hospital level | .010[a] | .016[a] | .168[a] | .037[a] | .108[a] | .100[a] | .207[a] | — |

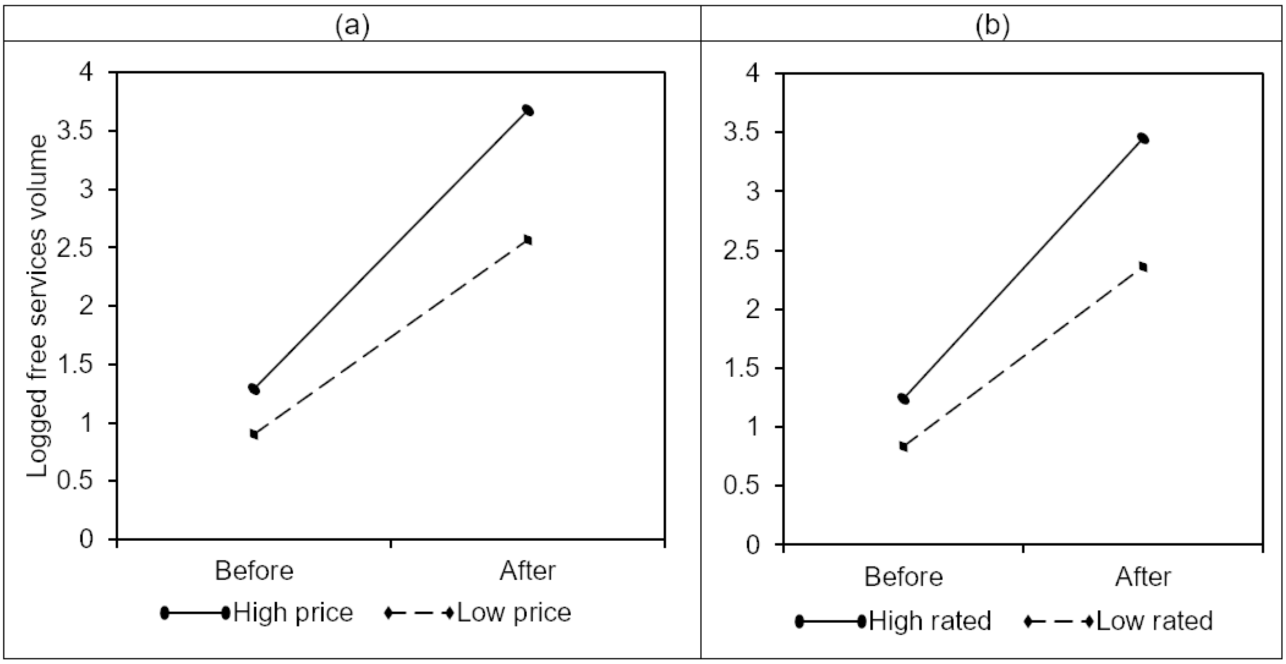
[a]$P<.01$.

## Results

### Model-Free Evidence

We compared the intensity of physicians’ online prosocial behaviors before and after the introduction of monetary incentives. To illustrate the moderation effects of high price and high rated, we generated a set of plots using the physicians in our data who had experienced monetary incentives (ie, treatment physicians). Figure 3 shows changes in the physicians’ online prosocial behavior after the introduction of monetary incentives by the level of physicians’ pricing strategy and online rating. The y-axis is the logged volume of free services offered by physicians. The amount of online prosocial behavior for both high self-recognition (high price) and low self-recognition (low price) physicians increased after the introduction of monetary incentives; the increase for high self-recognition physicians was much higher than for low self-recognition physicians. Similarly, the online prosocial behavior volume increased more for high-rated physicians than low-rated physicians. These patterns are consistent with the monetary incentive effects predicted in our three hypotheses and provide preliminary support to these hypotheses.

**Figure 3.** Model-free comparison of free services volume for physicians by (a) price setting and (b) rating.




## Model Estimation

We report the estimated effects of monetary incentives on physicians' online prosocial behavior in Table 3. First, we employed the ordinary least squares and the fixed-effects specifications to estimate the main effect of monetary incentives. Both specifications showed consistent results: the introduction of monetary incentives led to a significant increase in physicians' online prosocial behaviors, supporting hypothesis 1. Monetary incentives increased physicians' concerns about their prosocial self-image; they wanted to make more prosocial contributions to strengthen their online images. If the online image-strengthening effect held, there should be a stronger effect of introductory monetary incentives for physicians with high levels of self-recognition and others' recognition. Estimates with self-recognition and others recognition confirm our conjecture; therefore, hypotheses 2 and 3 are supported. The online image-strengthening effect provided by introductory monetary incentives worked better, especially for physicians with high image demands. The combined model also affirmed our hypotheses.

## Robustness Check

Although we controlled a set of observable attributes along with the physicians' and time-fixed effects, a potential problem with our DID design was that physicians are different. A particular concern was that physicians who enrolled later may have been attracted by the introductory incentive policy, or there was a different trend that caused a selection bias in our estimation.

To check the robustness of our research model, we used matching methods to select similar physicians from our control and treatment groups to replicate the main analyses shown in Table 4 [40,41]. Specifically, we used the physicians' personal characteristics (ie, seniority, hospital ranking, and registration date) and contribution characteristics (ie, gifts and letters of thanks received, total patients replied) to match the physicians. Finally, there were 5131 physicians matched by a caliper match (a caliper of 0.001). Table 4 presents the estimated results of the robustness check, which confirmed our findings of monetary incentives on physicians' online prosocial behavior. Therefore, the results of the robustness check are consistent with the results of the main model.

**Table 3.** Estimation results of the impact of monetary incentives on physicians' online prosocial behavior (N=777,110).[a]

| Variable | Prosocial behavior | | | | |
|---|---|---|---|---|---|
| | Ordinary least squares | Fixed effects | Self-recognition | Recognition from others | Combined model |
| Monetary incentive | 1.498[b] (0.019) | 1.057[b] (0.019) | 0.959[b] (0.022) | 0.876[b] (0.026) | 0.842[b] (0.026) |
| Monetary incentive × high price | —[c] | — | 0.275[b] (0.037) | — | 0.188[b] (0.040) |
| Monetary incentive × high rated | — | — | — | 0.325[b] (0.034) | 0.267[b] (0.036) |
| Control variables | | | | | |
| Letters | 1.259[b] (0.011) | 1.167[b] (0.011) | 1.152[b] (0.011) | 1.143[b] (0.011) | 1.137[b] (0.011) |
| Votes | 0.345[b] (0.008) | — | — | — | — |
| High price | 0.071[b] (0.023) | — | — | — | — |
| Title | −0.129[b] (0.009) | — | — | — | — |
| High rated | −0.167[b] (0.021) | — | — | — | — |
| High hospital level | −0.190[b] (0.019) | — | — | — | — |
| Physician fixed effects | No | Yes | Yes | Yes | Yes |
| Month trends | Yes | Yes | Yes | Yes | Yes |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | .413 | .230 | .231 | .232 | .232 |

[a]Robust standard errors are in parentheses.

[b]$P$<.01.

[c]Not applicable.

**Table 4.** Estimation results of robustness check using matched samples (N=425,469).[a]

| Variable | Prosocial behavior | | | | |
|---|---|---|---|---|---|
| | Ordinary least squares | Fixed effects | Self-recognition | Recognition from others | Combined model |
| Monetary incentive | 1.622[b] (0.032) | 0.989[b] (0.029) | 0.853[b] (0.034) | 0.767[b] (0.042) | 0.709[b] (0.042) |
| Monetary incentive × high price | —[c] | — | 0.319[b] (0.051) | — | 0.242[b] (0.054) |
| Monetary incentive × high rated | — | — | — | 0.349[b] (0.050) | 0.279[b] (0.053) |
| Control variables | — | — | — | — | — |
| Letters | 1.186[b] (0.016) | 1.157[b] (0.018) | 1.132[b] (0.018) | 1.125[b] (0.018) | 1.112[b] (0.018) |
| Votes | 0.401[b] (0.013) | — | — | — | — |
| High price | 0.041 (0.030) | — | — | — | — |
| Title | −0.018 (0.021) | — | — | — | — |
| High rated | −0.192[b] (0.029) | — | — | — | — |
| High hospital level | −0.101[b] (0.030) | — | — | — | — |
| Physician fixed effects | No | Yes | Yes | Yes | Yes |
| Month trends | Yes | Yes | Yes | Yes | Yes |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | .403 | .223 | .225 | .225 | .226 |

[a]Robust standard errors are in parentheses.

[b]$P$<.01.

[c]Not applicable.

## Discussion

### Summary of Findings

This study investigated the influence of monetary incentives on physicians' online prosocial behaviors. Based on self-determination theory, we developed three research hypotheses and established an empirical model based on a DID design. The results of our research model support our hypotheses. Accordingly, this work provides three key findings. First, we found that the introduction of monetary incentives has a positive effect on physicians' online prosocial behavior (as measured by free services offered). Second, this promotion effect is enhanced by physicians' self-recognition of their personal medical competence (as measured by higher price setting). Third, the extent of patients' recognition of physicians' medical competence (as measured by higher rating) also can strengthen the positive effect of introductory monetary incentives on physicians' online prosocial behaviors.

### Discussion of Research Results

Prosocial behavior refers to any behavior that is beneficial to others and society. Prior studies mainly focus on contribution to charity [12,21], medical treatment in hospital [20], endowment of money [42], volunteerism for the American Red Cross [26], and blood donations [13]. Recently, some researchers investigated a broader range of prosocial behavior types, such as individuals' knowledge-sharing behavior [43] and content contribution in social media [14,30]. In telemedicine markets, physicians provide free online medical consulting services, which are a type of prosocial behavior through the internet. However, little research has investigated physicians' prosocial behaviors in telemedicine markets. Our study addressed this gap based on self-determination theory and found a new factor (ie, introduction of monetary incentives) that significantly affects physicians' online prosocial behaviors.

Moreover, monetary incentives are often used to encourage contributors to improve their prosocial behaviors [13]. It is one of the key external incentives of prosocial performance. However, several studies have found that monetary incentives may backfire [42,44], and some researchers argue that rewards will introduce image concerns about appearing "greedy" instead of "prosocial" [12,13,16,26]. However, the effect of monetary incentives on prosocial behavior is a joint function of internal psychological processes and environmental factors. In telemedicine markets, online medical feedback is a repeatable behavior, meaning that physicians can contribute more free feedback to maintain and compensate for their prosocial images after accepting monetary rewards. Thus, we find that monetary incentives have a positive effect on the intensity of physicians' online prosocial behaviors, which provides new evidence against the backfire of monetary incentives in an online health care context.

In addition, the introduction of monetary incentives may not be equally important for all physicians, and their differences should be taken into consideration. According to self-determination theory, the satisfaction of competence makes them more likely to engage in prosocial behaviors. As competence is evaluated by oneself and others, satisfaction of competence is based on self-recognition and others' recognition of a user's ability and performance. In particular, if physicians obtain higher self-recognition and recognition from others of their competence, they will have higher satisfaction of competence. Therefore, they will be more likely to contribute free feedback and behave prosocially. Based on the previous discussion, we found that the extent of self-recognition and patients' recognition on physicians' competence can strengthen the positive effect of introductory monetary incentives on physicians' online prosocial behaviors, which gives us a better understanding of the mechanisms behind these behaviors.

### Limitations and Future Research

There are several limitations of this study. First, our observations are of only Chinese physicians. The incentive effects may differ in other countries due to cultural differences. Future studies can investigate this issue by leveraging cross-platform datasets. Second, this preliminary study investigates the general effects of incentives on physicians' online prosocial behaviors. More detailed settings, such as comparing online and offline environments, should be applied in future studies.

### Contributions

This study contributes to the literature in several ways. First, to our knowledge, this is the first study to investigate physicians' prosocial behaviors in the telemedicine context, which adds to both streams of eHealth and prosocial behavior. Although abundant studies have examined prosocial behaviors in offline markets [12,13,26], few have explored it in online markets, especially in an online health care context. In filling this research gap, our research extends the current understanding of online prosocial behavior through the consideration of free online medical consulting services offered by physicians. Second, we enrich the existing literature of factors that affect online prosocial behavior [26,30,32]. This research examined whether and how the introduction of a monetary incentive affects the online prosocial performance of physicians, which extends the current studies on online prosocial performance and related influence factors. The results confirm the effects of introductory monetary incentives on physicians' online prosocial performances and encourage future studies to consider it as an important perspective when studying online prosocial behavior. Third, this study deepens the literature of online prosocial behavior in specific mechanisms by considering the extent of recognition on physicians' competence [2,43]. We found that the promotion effect of monetary incentives on physicians' online prosocial behaviors is enhanced by physicians' self-recognition, and patients' recognition of physicians' medical competence. This study illuminates that offering online prosocial behaviors is an effective way to present real quality information and build reputation for physicians, which is an important insight in the existing literature of both marketing and eHealth. This provides novel insights into the future studies that tend to take specific business processes into account when studying online health care.

This study also offers some practical implications. First, this study indicates an effective approach to increase physicians' online prosocial behaviors by introducing monetary incentives. This can prompt physicians to improve their allocations of service. Ultimately, patients benefit more from these extra online prosocial behaviors. Second, our findings will shed light on the facilitating roles of physician traits by testing several practice-oriented variables (price setting and rating value), which provide valuable implications to practitioners. Physicians of different types can take corresponding measures to promote themselves by providing prosocial behaviors in online platforms. Due to the imbalance issue of increasing online medical demands and limited eHealth system resources globally, physicians' online prosocial behaviors are effective ways to compensate for medical services online and offline.

### Conclusions

Our study investigates physicians' online prosocial behaviors through self-determination theory embedded in an online health care platform. We extend self-determination theory in the online health care context and demonstrate the relationship between incentive mechanisms and the prosocial behaviors of physicians. The preliminary results support our theory-based model. We found that the introduction of monetary incentives has a positive effect on the volume of physicians' online prosocial behaviors, and the extent of self-recognition and others' recognition of physicians' competence can strengthen this promotion effect. This means that physicians with high self-recognition and others' recognition will make more prosocial contributions in online health care platforms.

## Acknowledgments

This study was supported by the National Social Science Foundation of China (14CXW045). The authors greatly appreciate the work of the editors and anonymous reviewers; we are grateful for their insightful comments and suggestions. All errors remain ours.

## Conflicts of Interest

None declared.

## References

1. Venkatesh V, Zhang X, Sykes TA. Doctors do too little technology: a longitudinal field study of an electronic healthcare system implementation. Inform Syst Rese 2011 Sep;22(3):523-546. [doi: 10.1287/isre.1110.0383]
2. Zhang X, Gu B, Guo X. The impact of introductory incentives on new users: evidence from an online health community. 2017 Presented at: International Conference on Information Systems (ICIS); Dec 10-13, 2017; Seoul, Korea.
3. Zhang X, Guo X, Lai K, Yi W. How does online interactional unfairness matter for patient–doctor relationship quality in online health consultation? The contingencies of professional seniority and disease severity. Eur J Inform Syst 2018 Nov 27;28(3):336-354. [doi: 10.1080/0960085X.2018.1547354]
4. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. J Manage Inform Syst 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
5. Meng F, Guo X, Peng Z, Lai K, Zhao X. Investigating the adoption of mobile health services by elderly users: trust transfer model and survey study. JMIR Mhealth Uhealth 2019 Jan 08;7(1):e12269 [FREE Full text] [doi: 10.2196/12269] [Medline: 30622092]
6. Yang H, Zhang X. Investigating the effect of paid and free feedback about physicians' telemedicine services on patients' and physicians' behaviors: panel data analysis. J Med Internet Res 2019 Mar 22;21(3):e12156 [FREE Full text] [doi: 10.2196/12156] [Medline: 30900997]
7. Guo F, Zou B, Guo J, Shi Y, Bo Q, Shi L. What determines academic entrepreneurship success? A social identity perspective. Int Entrep Manag J 2019 Mar 15:1-24. [doi: 10.1007/s11365-019-00569-6]
8. Yang H, Guo X, Wu T, Ju X. Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision. Electron Commer R A 2015 May;14(3):192-203. [doi: 10.1016/j.elerap.2015.04.001]
9. Yaraghi N, Wang W, Gao GG, Agarwal R. How online quality ratings influence patients' choice of medical providers: controlled experimental survey study. J Med Internet Res 2018 Dec 26;20(3):e99 [FREE Full text] [doi: 10.2196/jmir.8986] [Medline: 29581091]
10. Deci EL, Ryan RM. The what and why of goal pursuits: Human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
11. Benabou R, Tirole J. Intrinsic and extrinsic motivation. Rev Econ Stud 2003 Jul;70(3):489-520. [doi: 10.1111/1467-937X.00253]
12. Ariely D, Bracha A, Meier S. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. Am Econ Rev 2009 Mar;99(1):544-555. [doi: 10.1257/aer.99.1.544]
13. Lacetera N, Macis M. Do all material incentives for pro-social activities backfire? The response to cash and non-cash incentives for blood donations. J Econ Psychol 2010 Aug;31(4):738-748. [doi: 10.1016/j.joep.2010.05.007]
14. Tang Q, Gu B, Whinston AB. Content contribution for revenue sharing and reputation in social media: a dynamic structural model. J Manage Inform Syst 2014 Dec 08;29(2):41-76. [doi: 10.2753/MIS0742-1222290203]

15. Hendijani R, Bischak DP, Arvai J, Dugar S. Intrinsic motivation, external reward, and their effect on overall motivation and performance. Hum Perform 2016 Apr 29;29(4):251-274. [doi: 10.1080/08959285.2016.1157595]
16. Bénabou R, Tirole J. Incentives and prosocial behavior. Am Econ Rev 2006 Dec;96(5):1652-1678. [doi: 10.1257/aer.96.5.1652]
17. Bagozzi RP, Moore DJ. Public service advertisements: emotions and empathy guide prosocial behavior. J Marketing 1994 Jan;58(1):56-70. [doi: 10.1177/002224299405800105]
18. Karakashian LM, Walter MI, Christopher AN, Lucas T. Fear of negative evaluation affects helping behavior: the bystander effect revisited. N Am J Psychol 2006;8(1):13-32 [FREE Full text]
19. Batson CD, Powell AA. Altruism prosocial behavior. Handb Psychol 2003 Apr:463-484. [doi: 10.1002/0471264385.wei0519]
20. Carmel S, Glick SM. Compassionate physicians: Personality traits and prosocial attitudes. Psychol Rep 2016 Dec 06;73(3_suppl):1362-1362. [Medline: 8115589]
21. Furse DH, Stewart DW. Monetary incentives versus promised contribution to charity: New evidence on mail survey response. J Marketing Res 1982 Aug;19(3):375-380. [doi: 10.1177/002224378201900311]
22. Kreps DM. Intrinsic motivation and extrinsic incentives. The American Economic Review 1997 May;87(2):359-364.
23. Deci EL. Effects of externally mediated rewards on intrinsic motivation. J Pers Soc Psychol 1971;18(1):105-115. [doi: 10.1037/h0030644]
24. Gneezy U, Rustichini A. Pay enough or don't pay at all. Q J Econ 2000 Aug;115(3):791-810. [doi: 10.1162/003355300554917]
25. Frey BS, Jegen R. Motivation crowding theory. J Econ Surv 2002 Dec 16;15(5):589-611. [doi: 10.1111/1467-6419.00150]
26. Exley C. Incentives for prosocial behavior: the role of reputations. Manage Sci 2018 May;64(5):2460-2471. [doi: 10.1287/mnsc.2016.2685]
27. Manski CF. Economic analysis of social interactions. J Econ Perspect 2000 Aug;14(3):115-136. [doi: 10.1257/jep.14.3.115]
28. Gneezy U, Meier S, Rey-Biel P. When and why incentives (don't) work to modify behavior. J Econ Perspect 2011 Nov;25(4):191-210. [doi: 10.1257/jep.25.4.191]
29. Grossman Z, van der Weele JJ. Self-image and willful ignorance in social decisions. J Eur Econ Assoc 2017;15(1):173-217. [doi: 10.1093/jeea/jvw001]
30. Toubia O, Stephen AT. Intrinsic vs. image-related utility in social media: why do people contribute content to Twitter? Market Sci 2013 May;32(3):368-392. [doi: 10.1287/mksc.2013.0773]
31. Ellingsen T, Johannesson M. Pride and prejudice: the human side of incentive theory. Am Econ Rev 2008 May;98(3):990-1008. [doi: 10.1257/aer.98.3.990]
32. DellaVigna S, List JA, Malmendier U. Testing for altruism and social pressure in charitable giving. Q J Econ 2012 Jan 18;127(1):1-56. [doi: 10.1093/qje/qjr050]
33. Wu H, Lu N. How your colleagues' reputation impact your patients' odds of posting experiences: evidence from an online health community. Electron Commer R A 2016 Mar;16:7-17. [doi: 10.1016/j.elerap.2016.01.002]
34. 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]
35. Marylène G, Muraleedharan LD. Self-determination theory and work motivation. J Organ Behav 2005 Apr;26(4):331-362. [doi: 10.1002/job.322]
36. Gagné M. The role of autonomy support and autonomy orientation in prosocial behavior engagement. Motiv Emotion 2003 Sep;27(3):199-223. [doi: 10.1023/A:1025007614869]
37. Angrist JD, Pischke JS. Mostly harmless econometrics: an empiricist's companion. In: Business Economics. Princeton, NJ: Princeton University Press; 2008.
38. Bertrand M, Duflo E, Mullainathan S. How much should we trust differences-in-differences estimates? Q J Econ 2004 Feb 01;119(1):249-275. [doi: 10.1162/003355304772839588]
39. Greenwood BN, Wattal S. Show me the way to go home: an empirical investigation of ride sharing and alcohol related motor vehicle homicide. MIS Quart 2017 Jan 1;41(1):163-187. [doi: 10.25300/MISQ/2017/41.1.08]
40. Abadie A, Imbens GW. Large sample properties of matching estimators for average treatment effects. Econometrica 2006 Jan;74(1):235-267. [doi: 10.1111/j.1468-0262.2006.00655.x]
41. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. J Econ Surv 2008 Feb;22(1):31-72. [doi: 10.1111/j.1467-6419.2007.00527.x]
42. Masclet D, Noussair C, Tucker S, Villeval M. Monetary and nonmonetary punishment in the voluntary contributions mechanism. Am Econ Rev 2003 Feb;93(1):366-380. [doi: 10.1257/000282803321455359]
43. Hung SY, Durcikova A, Lai HM, Lin WM. The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior. Int J Hum-Comput Stud 2011 Jun;69(6):415-427. [doi: 10.1016/j.ijhcs.2011.02.004]
44. Meier S, Stutzer A. Is volunteering rewarding in itself? Economica 2007 Sep 24;75(297):39-59. [doi: 10.1111/j.1468-0335.2007.00597.x]


## Abbreviations

**DID:** difference in differences

*Edited by G Eysenbach; submitted 12.05.19; peer-reviewed by P Luo, E Zhang; comments to author 06.06.19; revised version received 18.06.19; accepted 27.06.19; published 26.07.19.*

*Please cite as:*
*Jing D, Jin Y, Liu J*
*The Impact of Monetary Incentives on Physician Prosocial Behavior in Online Medical Consulting Platforms: Evidence From China*
*J Med Internet Res 2019;21(7):e14685*
*URL: http://www.jmir.org/2019/7/e14685/*
*doi:10.2196/14685*
*PMID:31350834*

Original Paper

# Overcoming Barriers to Mobilizing Collective Intelligence in Research: Qualitative Study of Researchers With Experience of Collective Intelligence

Van Thu Nguyen[1,2,3], PharmB, MPH; Bridget Young[3], PhD; Philippe Ravaud[1,2,4], MD, PhD; Nivantha Naidoo[1], MPH, MD; Mehdi Benchoufi[1,2,4], MD, PhD; Isabelle Boutron[1,2,4], MD, PhD

[1]INSERM, U1153 Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (CRESS), Methods of Therapeutic Evaluation of Chronic Diseases Team (METHODS), Paris, France

[2]University Paris Descartes, Paris, France

[3]Department of Health Services Research, Institute of Population Health Sciences, University of Liverpool, Liverpool, United Kingdom

[4]Centre d'Epidémiologie Clinique, Hôpital Hôtel Dieu, Assistance Publique des Hôpitaux de Paris, Paris, France

**Corresponding Author:**
Van Thu Nguyen, PharmB, MPH
INSERM
U1153 Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (CRESS)
Methods of Therapeutic Evaluation of Chronic Diseases Team (METHODS)
Paris,
France
Phone: 33 142347825
Email: van.nguyen@clinicalepidemio.fr

## Abstract

**Background:** Innovative ways of planning and conducting research have emerged recently, based on the concept of collective intelligence. Collective intelligence is defined as shared intelligence emerging when people are mobilized within or outside an organization to work on a specific task that could result in more innovative outcomes than those when individuals work alone. Crowdsourcing is defined as "the act of taking a job traditionally performed by a designated agent and outsourcing it to an undefined, generally large group of people in the form of an open call."

**Objective:** This qualitative study aimed to identify the barriers to mobilizing collective intelligence and ways to overcome these barriers and provide good practice advice for planning and conducting collective intelligence projects across different research disciplines.

**Methods:** We conducted a multinational online open-ended question survey and semistructured audio-recorded interviews with a purposive sample of researchers who had experience in running collective intelligence projects. The questionnaires had an interactive component, enabling respondents to rate and comment on the advice of their fellow respondents. Data were analyzed thematically, drawing on the framework method.

**Results:** A total of 82 respondents from various research fields participated in the survey (n=65) or interview (n=17). The main barriers identified were the lack of evidence-based guidelines for implementing collective intelligence, complexity in recruiting and engaging the community, and difficulties in disseminating the results of collective intelligence projects. We drew on respondents' experience to provide tips and good practice advice for governance, planning, and conducting collective intelligence projects. Respondents particularly suggested establishing a diverse coordination team to plan and manage collective intelligence projects and setting up common rules of governance for participants in projects. In project planning, respondents provided advice on identifying research problems that could be answered by collective intelligence and identifying communities of participants. They shared tips on preparing the task and interface and organizing communication activities to recruit and engage participants.

**Conclusions:** Mobilizing collective intelligence through crowdsourcing is an innovative method to increase research efficiency, although there are several barriers to its implementation. We present good practice advice from researchers with experience of collective intelligence across different disciplines to overcome barriers to mobilizing collective intelligence.

***(J Med Internet Res 2019;21(7):e13792)*** doi:10.2196/13792

**KEYWORDS**

collective intelligence; crowdsourcing; open innovation; health; research; survey; interview

## Introduction

Innovative ways of conducting research have emerged recently with promising results. For example, Harvard Medical School organized an ideas competition, which attracted participants from 17 countries who contributed 150 new research ideas for managing type 1 diabetes [1]. In China, a creative competition involving participants from diverse backgrounds such as graphic designers, artists, and students resulted in new interventions to increase the HIV testing rate [2]. These initiatives were based on methods of mobilizing collective intelligence through crowdsourcing [3,4]. Collective intelligence is defined as shared intelligence emerging when people are mobilized within or outside an organization to work on a specific task that could result in more innovative outcomes than those when individuals work alone [5]. Crowdsourcing is "the act of taking a job traditionally performed by a designated agent and outsourcing it to an undefined, generally large group of people in the form of an open call" [6]. Although collective intelligence can emerge from day-to-day collaboration in science, by crowdsourcing, a large number of individuals with diverse backgrounds and expertise are enabled to contribute to research, resulting in collective intelligence on a large scale [3].

Use of such methods is increasing markedly across different disciplines. From 2010 to 2018, the number of projects mobilizing collective intelligence hosted by the US government through the website www.challenge.gov increased by more than 250% [7]. Collective intelligence enables researchers to solve problems, generate new research ideas, create intellectual products, and critically appraise research ideas and work [8-15]. For example, an initiative called Transparence Science created an online community of physicians and patients to develop a clinical trial protocol together [16].

Some resources describing methods of mobilizing collective intelligence in health research have been published [17,18]. However, literature on barriers that researchers encounter across different disciplines when mobilizing collective intelligence, advice on how to overcome these barriers, and good practice in mobilizing collective intelligence is still lacking. Our study aimed to identify the barriers to mobilizing collective intelligence and ways to overcome these barriers and provide good practice advice for those planning and conducting collective intelligence projects across different disciplines.

## Methods

### Study Design

To investigate collective intelligence methods, we conducted (1) a multinational online open-ended survey that allowed us to access the perspectives of a diverse group of respondents involved in collective intelligence and (2) semistructured interviews that allowed for more in-depth exploration of respondents' perspectives on this fairly new topic.

Our approach was pragmatic when providing insights on the methods of mobilizing collective intelligence, but interpretive when analyzing respondents' reports as subjective accounts of their experience when using these methods. The study received ethical approval (Ref: 17-386) from French National Institute of Health and Medical Research Ethic Committee (IRB00003888).

### Reflexivity

We have extensive experience in clinical trial methodology and an interest in understanding the method of mobilizing collective intelligence through crowdsourcing to apply it in clinical research. Some members of the team have conducted projects mobilizing collective intelligence.

### Sample and Recruitment

We recruited principal investigators and project coordinators experienced in running collective intelligence projects. We purposively sampled these researchers, seeking diversity in terms of their experience of different collective intelligence methods and their disciplinary backgrounds. We identified authors of articles reporting a project using collective intelligence [8], included researchers in the network of European citizen science association [19], and invited speakers in collective intelligence conferences [20,21]. We also used snowball sampling, asking respondents to send us email addresses of colleagues active in the field of collective intelligence.

An invitation email was sent via Mailjet [22] to researchers and project coordinators whose email addresses were available. The invitation contained a link to the first page of the survey, through which they indicated their consent. Two reminder emails were sent to nonrespondents.

We proposed semistructured interviews to a purposive sample of 24 researchers who did not respond to the first email invitation and who were mainly using collective intelligence in biomedical research and citizen science projects. They were invited via a personalized email sent by VN.

### Online Open-Ended Survey

The survey was developed using the results of a scoping review [8] and then pilot tested (Multimedia Appendix 1). It comprised five closed-ended questions to identify respondents' background and expertise, and four open-ended questions exploring their motivation and experience with mobilizing collective intelligence, particularly the barriers they encountered and their solutions (Textbox 1). Finally, respondents were asked to provide three pieces of advice to a colleague planning to use collective intelligence in a project for the first time. To promote interaction between participants, we also asked them to rate and comment on the advice that other respondents had entered; the advice shown to each respondent was randomly selected from the pool of advice provided by previous respondents.

**Textbox 1.** Open-ended questions in the online survey.

- What are *the benefits of collective intelligence* that made you decide to use it in your project?
- What were the *most important factors* contributing to the success of mobilizing collective intelligence in your project?
- What were the *most challenging issues* you had to face when using collective intelligence in your project and *your solutions for those challenges* (eg, difficulties in identifying and motivating participants, designing tasks for participants, evaluate quality of participants' contribution, decision making)?
- What *three pieces of advice* would you give to a colleague who intends to use collective intelligence in a project for the first time?
- Please read the advice from another participant. (Showing an answer from another participant). What do you think of this advice?

### Semistructured Interviews

We sent individuals who expressed an interest in being interviewed an information sheet about the study. Interviews were conducted according to participants' convenience (eg, face-to-face, telephone, and teleconference [gotomeeting.com]), and oral consent was obtained.

The interview guide covered content similar to that of the survey questionnaire (Multimedia Appendix 2). VN conducted all interviews in English. These were audio recorded, transcribed verbatim by a native English-speaking transcriber, and anonymized by VN. Interviews lasted between 22 minutes and 1 hour (median: 34 minutes). After each interview, VN wrote a summary of the interview to record the reflections on the interview and initial thoughts for the analysis.

### Analysis

Analysis of open-ended survey responses and interview transcripts was thematic, drawing on the framework analysis [23,24]. The analysis was partly deductive, with some aspects informed by the previous literature on collective intelligence, but also inductive to identify new themes and ensure that the analysis was grounded in the data. VN led the analysis. Two senior researchers BY and IB periodically reviewed transcripts and commented on the developing analysis.

Open codes and categories were developed by a constant comparative approach, reading and re-reading data and considering it in the context of other data from the same respondent and in the context of the wider dataset [25]. An initial framework of themes and subthemes was developed based on the first eight interview transcripts and then imported into NVivo to code the remaining transcripts and survey entries. The framework was further refined throughout the process of analysis.

Data saturation was examined by the theme accumulation curve that presented the number of distinct themes generated against a number of units of analysis used to generate those distinct themes (Multimedia Appendix 3) [26].

Respondents' survey comments on the advice provided by other respondents were categorized as "agree" (ie, positive comments), "disagree" (ie, negative comments), and "neither agree nor disagree" (ie, neutral comment or did not directly comment on the idea in the answer). Two coders (VN and NN) independently assessed the content of each comment and discussed this to reach consensus. We received 129 pieces of advice: 100 advices were commented on by other respondents, and 28 were commented on twice, resulting in 128 comments. Most comments (98/128, 77%) agreed with the advice provided by respondents, and only 9% (12/128) disagreed. We summarized advice that commentators disagreed with in Multimedia Appendix 4.

The themes described below are derived from both interviews and survey entries. We present excerpts from the interviews and survey to explicate the findings and our interpretation of the data. Interviewees are indicated by "I" and survey respondents are indicated by "S"; "[…]" denotes text removed for brevity. Research disciplines of interviewees and survey respondents are listed in Multimedia Appendix 5.

### Data Sharing

The anonymized data from the online survey will be deposited on Zenodo, an open-access research data repository. Anonymized transcripts of interviews will be provided upon request.

## Results

### Respondent Characteristics

Of 157 people who clicked the survey link, 65 participated in the survey. Of the 24 people who were invited for interview, 17 participated in it. Of those who were not interviewed, two were unable to schedule an interview within the time frame of the study, two advised the interviewer to contact another team member responsible for the projects, two did not respond, and one was unable to be interviewed in English. Table 1 presents the demographic characteristics of survey respondents and interviewees. Survey participants were mainly from the field of computer science (43%), while interviewees were mainly involved in biomedicine and health care (59%). They mostly mobilized collective intelligence to solve research problems (70%) and generate new ideas (46%).

**Table 1.** Respondent demographics.

| Demographic information | Survey respondents (N=65)[a], n (%) | Interviewees (N=17), n (%) | Total (N=82), n (%) |
|---|---|---|---|
| **Age groups (years)** | | | |
| 20-29 | 4 (6) | 0 (0) | 4 (5) |
| 30-39 | 27 (42) | 1 (6) | 28 (34) |
| 40-49 | 19 (30) | 11 (65) | 30 (37) |
| 50-59 | 8 (12) | 3 (18) | 11 (13) |
| ≥60 | 4 (6) | 2 (12) | 6 (7) |
| **Location** | | | |
| Europe | 42 (65) | 11 (65) | 53 (65) |
| North America | 18 (28) | 6 (35) | 24 (29) |
| Asia | 2 (3) | 0 (0) | 2 (2) |
| **Research field[b]** | | | |
| Computer science | 28 (43) | 2 (12) | 30 (37) |
| Biomedicine and health care | 9 (14) | 10 (59) | 19 (23) |
| Engineering and technology development | 9 (14) | 0 (0) | 9 (11) |
| Economics, commercial, and business development | 7 (11) | 2 (12) | 9 (11) |
| Education and information studies | 7 (11) | 0 (0) | 7 (9) |
| Environmental science | 5 (8) | 2 (12) | 7 (9) |
| Psychology and social science | 5 (8) | 0 (0) | 5 (6) |
| Laws, politics, and governance | 4 (6) | 1 (6) | 5 (6) |
| Other | 10 (15) | 0 (0) | 10 (12) |
| **Purpose of using collective intelligence in their projects[b]** | | | |
| Solve problems (ie, participants propose solutions to difficulties given by organizers) | 44 (68) | 13 (76) | 57 (70) |
| Generate ideas (ie, participants contribute to new ideas for research and development) | 32 (49) | 6 (35) | 38 (46) |
| Evaluate ideas (ie, participants evaluate the quality of the ideas/work) | 23 (35) | 1 (6) | 24 (29) |
| Create intellectual outputs (ie, participants create health education materials, clinical trial protocols, and prognostic models) | 16 (25) | 1 (6) | 17 (21) |
| Other | 10 (15) | 0 (0) | 10 (12) |

[a]Data for two persons are missing.

[b]Respondents selected more than one option.

## Researchers' Motivations for Mobilizing Collective Intelligence

Participants reported trying the methods of collective intelligence as a new way of conducting research because traditional research methods no longer fitted their needs (Table 2). They commented that research questions were becoming very complex, unlikely to be solved within a single discipline and by traditional models of research, where each team working in relative isolation impeded research efficiency.

Respondents also noted the personal "pleasure" they derived from working "in teams with other people" (I10). Collective intelligence helped make research more enjoyable and helped them "to find some bridge, to…better understand each other, work closely together and this has some huge impact." (I02)

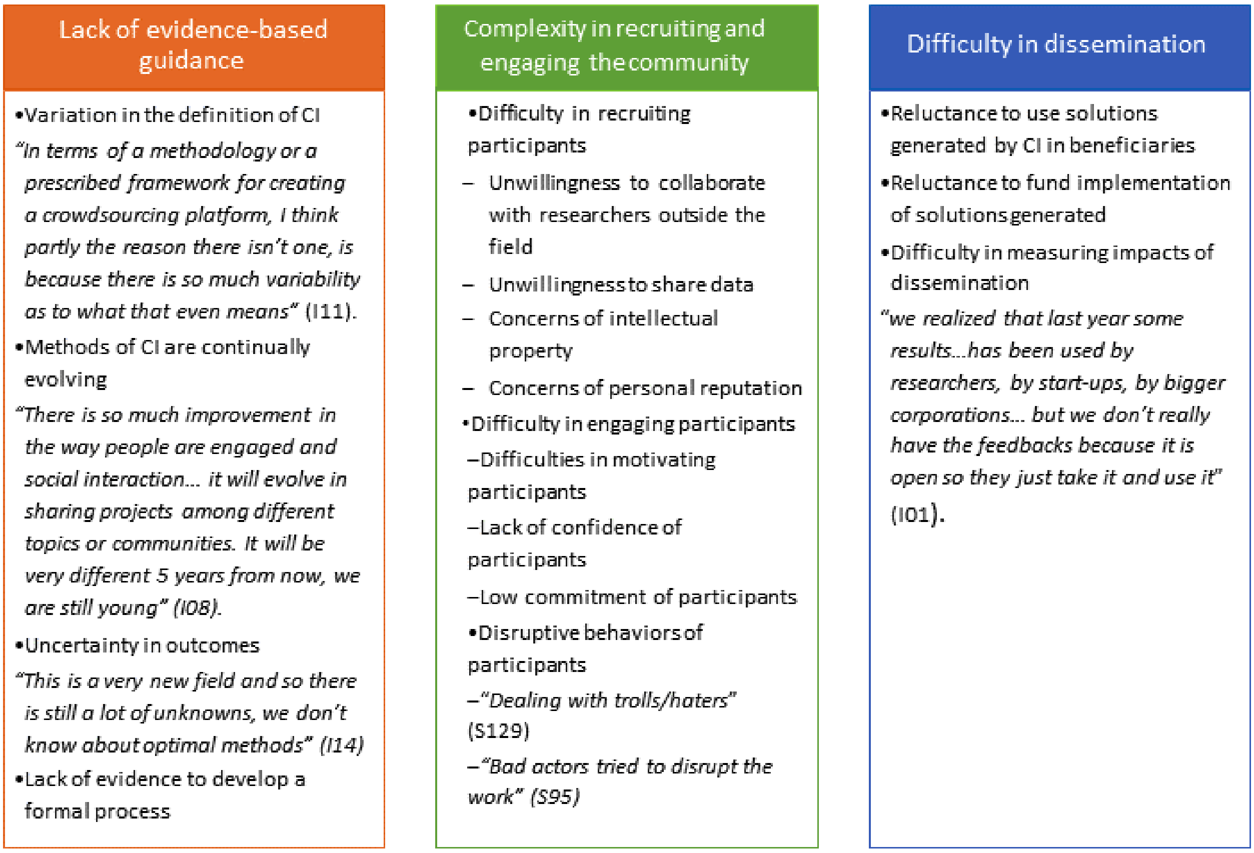
## Barriers To Mobilizing Collective Intelligence

Although collective intelligence has numerous benefits, respondents found aspects of collective intelligence challenging. These challenges, in part, arose from the novelty of the method and complexity in engaging the community (Figure 1).

**Table 2.** Reasons for mobilizing collective intelligence.

| Issues with traditional research practice | How collective intelligence can address the issue |
|---|---|
| Research questions were becoming more complex, and the answers could not be found from a single discipline | Collective intelligence provided the opportunity to work with people with different types of expertise and integrate their skills to solve problems from different angles: *Knowledge is distributed in different domains and some "wicked" questions cannot be answered within a single discipline or sector, ie, we need both different science disciplines as well as expertise from the practice and policy sector.* (S75) |
| Current research was conducted inefficiently by "repeating efforts" (I06) | Collective intelligence allowed researchers to conduct research as collective efforts where different approaches to a research question could be collectively and thoroughly evaluated to avoid redundant efforts: *In science, often we are developing solutions independently and we are kind of repeating erm…efforts, […] an alternative is to post a problem or a question to the research community and then just see what kind of solutions people come up with, and possibly combine these solutions and that you could call CI.* (I06) |
| As research questions became more complex, conducting research required a longer time. Researchers would not have enough time to investigate different aspects. "It takes for hundreds of years…you will never [be able to] explore everything." (I08) | With a large community contributing, researchers were able to finish work within shorter time scales: *Draw on the experiences and expertise of a varied group of people to advance and implement ideas that would take a significantly longer time to solve as an individual.* (S104) |
| It was more costly to work with experts in the field and took longer to engage them | Mobilizing contribution from a wide community was cheaper than working with experts in the field, yet the former could achieve the same outcomes: *Our organization has done over 300 crowd-based challenges and has found success in 80-90% of those challenges with cost and schedule savings in the majority of them.* (S49) |

**Figure 1.** Barriers to mobilizing collective intelligence. CI: collective intelligence.



### *Lack of Evidence-Based Guidelines on Methods of Mobilizing Collective Intelligence*

Use of collective intelligence through crowdsourcing in research is relatively new. Some respondents reported that they had delved into this method before they had become fully aware of the concepts of collective intelligence, crowdsourcing, or citizen science. Respondents also recounted challenges they had faced in their projects due to lack of evidence for an "optimal method" (I14) and noted the absence of a methodological guide for them to follow.

### *Complexity in Recruiting and Engaging the Community of Participants*

Respondents believed that some potential collective intelligence participants had "a lot of prejudice" (I03) toward collaborating with people from different fields, and it was "not easy to make them to participate" (I02) in collective intelligence projects. One interviewee (I06) working in the field of biomedicine spoke of the difficulties he experienced in motivating industrial partners to work with academic institutions in his challenge contests. He commented that collective intelligence participants had concerns about the ownership of the research intellectual property of solutions created and negative reputation consequences if their solutions performed poorly.

Respondents described difficulties in "retaining all the people that sign up…to get them to actually participate" (I09), as most participants joined collective intelligence as a side project or "an unfunded kind of project" (I12). They also believed that many potential collective intelligence participants were "not confident enough" (I07), which hindered their contribution.

Respondents reported situations when participants had tried to cheat or behaved aggressively, which adversely influenced the community and demotivated other participants. One interviewee shared his experience with this disruptive behavior, when organizing challenge contests for data analytics*:*

> *They will make different identities…and…submit hundreds [times]…[they] cheat the leader boards. [They] will discourage many people from [participating]…but [they]don't have the solution.* [I04]

He explained that this disruptive behavior partly arose from the competitive nature of a contest, adding that participants might be under pressure from their organizations to win international contests in order to increase reputation of the organizations.

### Difficulties in Disseminating the Solutions Generated by Collective Intelligence

Respondents found it challenging to disseminate and implement the findings of their collective intelligence projects to the relevant communities, as funders and beneficiaries were unfamiliar with this emerging method. These challenges arose partly from the "prejudice" of researchers (I03) that people who were outside of the field might not have sufficient capacity to create solutions. One interviewee spoke of his difficulty in persuading funders to sponsor the further development of solutions generated by collective intelligence participants in a challenge contest that he had organized:

> *The third challenge…was getting people to recognize that these solutions existed and were available…there is a reluctance to use crowdsource and open source solutions like this.* [I15]

### Good Practice Advice for Planning and Conducting Collective Intelligence Projects

When describing their projects, respondents reflected on the solutions that they had considered or used to overcome these barriers. We also explicitly sought their advice on what they perceived to be good practice in planning and conducting collective intelligence projects. In the sections that follow, we present respondents' good practice recommendations for collective intelligence projects, covering three main themes: project governance, planning, and conduct of collective intelligence projects (Figure 2).

**Figure 2.** Good practice advice for planning and conducting research mobilizing collective intelligence.


## The Project Governance

### *Establishing a Coordination Team*

Respondents advised researchers to establish a coordination team dedicated to supporting projects mobilizing collective intelligence. They suggested that the coordination team should include people with diverse expertise to bring more “insights” (I01) to the project and help with “getting leadership and [funders] on board” (S23). Respondents also encouraged researchers to involve stakeholders and representatives from potential collective intelligence participant groups in planning, designing, and conducting collective intelligence projects.

> *Listen very carefully to your participants and work with them. Ensure mutual benefits in your design and co-create the project.* [S62]

Respondents advised that the involvement of participants’ representatives from early stage would help identify mutual research interests between participants and researchers, design appropriate tasks, and develop effective communication strategies to engage potential participants. Respondents also emphasized on the importance of including people with experience in communication in the team to support recruitment and engagement activities with collective intelligence participants.

### *Create a Set of Common Rules*

Respondents suggested that the coordination team create a set of common rules for collective intelligence participants to encourage mutual respect and constructive contributions. They mentioned the use of “diplomatic quality control” (I03) to flag aggressive or disruptive behavior from participants and to try to create a participatory and friendly environment for others to freely contribute their work. They also suggested preparing a resolution plan to resolve conflicts between collective intelligence participants.

## Planning a Collective Intelligence Project

### *Identify the Research Question*

Respondents commented that an early step in research involving collective intelligence was to identify “an interesting problem” (I06) with “high scientific value” (I04) that would gain from the involvement of a large and diverse community.

> *It is number one that there is a problem out there worth solving […], a project that it makes sense to try and bring in…people outside of the normal kind of scope or expertise area for it.* [I15]

They noted that identifying “just difficult enough” (I06) problems and “putting yourself in the participants’ [positions]” (I08) were crucial to create appropriate research problems to gain buy-in from target communities. One interviewee (I15) working in the field of biomedicine and health care described how a dynamic process involving “a lot of conversations” was part of the process of establishing whether the community would be interested in the research problem.

> *We knew there were a lot of people…working on it [the research topic] and no one had come up with an optimal solution and we felt like there were enough people who would be interested…but that didn’t come from us just sitting in a room alone. We actually reached out to many of the people…to see if they felt like there was a need and an ability to really take this further.*

### *Identify Communities of Participants*

Respondents also considered the choice of the communities a key factor in ensuring successful mobilizing of collective intelligence. Respondents suggested identifying communities who “have most contact with these problems” (I05).

> *You need to have champions of the cause…if you are doing something on Alzheimer’s, finding a person…who has Alzheimer’s, who their mother, father has Alzheimer’s and who has a personal vested interest and a strong…passion for the cause.* [I14]]

They emphasized on two important characteristics of the community—diversity and independence. Diversity in participants was thought to be important to generate novel solutions to the research problem. Diversity could be achieved by involving a larger number of participants with various disciplines.

> *The more participants you have, the more likely some of them will come with the new idea.* [I04]

Similarly, maintaining the independence of participants as they worked on the research problem was crucial to “free the minds and let [participants] think freely” (S104), allow “outside of the box thinking” (S146), and ensure that participants could voice their ideas without being influenced by a dominant opinion.

### *Decide on Incentives to Engage Participants*

Respondents suggested offering a combination of both extrinsic motivators such as authorship and access to the data and intrinsic motivators such as making tasks enjoyable, offering participants the opportunity to gain new knowledge and finding meaningful outlets for their skills. They described some innovative activities to engage participants:

> *Some of the things that we have done have been really fun, and really brought the community together…to create…a sense of community…like the 24-hour citation screening challenges. Where we have had hundreds of people, online at the same time, all with a specific target to try and reach within 24 hours…and those have been hugely exciting, really popular.* [I17]

Interestingly, some respondents tried to “avoid monetary prizes” (I14), as they believed that “the crowd may only be interested in the compensation and therefore, may take short-cuts or cheat if the task allows for that” (S153). Instead, they suggested offering research partnership, mentorship, or training as ways to benefit participants’ professional development.

### *Determine Methods to Evaluate Solutions Created by Collective Intelligence and Decision Making*

Respondents emphasized the need to “set up objective methods to validate the results” (S65), for example, by establishing a panel with diverse expertise to comprehensively evaluate contribution of participants. They also acknowledged the need to allow enough time for evaluation, given the large number of participants, and advised involving the crowd in the evaluation to increase the efficiency of the process. Automating screening of participants’ contributions was also suggested to reduce work load for the panel when performing the evaluation.

## Conducting Collective Intelligence Projects

### *Prepare Tasks and Interface*

Respondents highlighted the need to design a user-friendly interface to “make it really easy for people to contribute even if they have only got a minute free” (I17). They explained that “the design of the interfaces or platforms which people will use is often overlooked but can influence the results or the ease of data collection” (S25).

They also advised researchers to prepare training materials and offer tutorials to explain the project to participants and equip them with essential skills. However, they noted that the training should avoid providing participants with examples that could hinder participants’ creativity.

Respondents also recommended “verifying if it [the task and interface] works on small scale” (S16) and gradually scaling up. The pilot phase could help researchers foresee any technical and ethical issues related to data collection and participants’ identities, which could be addressed before a large number of collective intelligence participants enrolled.

### *Create a Clear Description of the Research Problem*

*Crafting* a clear description of the problem in a language relevant to those communities was considered a key step to helping collective intelligence participants understand the project objectives and judge whether they had the relevant skills to participate:

> *Good communication of a complex objective or complex data set…is not…always easy...if there is something that you don't even understand,…you won't put your time in that challenge* [I10]

One respondent also suggested dividing the objectives into concrete deliverables with clear requirements for participants' contributions:

> *In order for the collective to provide "intelligence" as opposed to noise, one must be very careful about what one measures… If the measures are ambiguous to the participants, or if there exists a short-cut for the participants to satisfy immediate goal without actually contributing to the overall big picture, many participants will find this short-cut and will explore it* [S20]

### Organize Communication Activities to Recruit Participants

Respondents described how they had organized various communication activities to recruit participants via advertisements on social media (eg, Google, Facebook, and other websites) and announcements in scientific publications. Several thought of working with an intermediary online platform, which had a readily available online community, as a practical approach for those who were new to collective intelligence. They advised researchers to partner with local organizations such as nongovernmental organizations, universities, and patient organizations and organize face-to-face meetings to connect directly with participants.

### Engage Participants Through Responsive Communication

To engage participants effectively, respondents believed that communicating frequently with collective intelligence participants, even being available *24/7* to guide them and give feedback on their contributions. Respondents believed this would improve the quality of participants' contributions and increase their commitment. Further, through responsive communication with participants, researchers could understand what resources participants needed for developing an implementable solution. Although virtual communication helped in ensuring responsive communication, respondents advised supplementing this with face-to-face engagement events to increase trust and create a sense of community among collective intelligence participants.

### Disseminate Solutions Created by Collective Intelligence for Beneficiaries and Collective Intelligence Participants

Respondents advised researchers to diversify the dissemination of their project findings through multiple channels and make the results open access to the public through social media.

Respondents suggested involving leaders of organizations from the beginning of the projects to ensure their support for implementation of solutions generated by collective intelligence. They encouraged other researchers using collective intelligence to "show their results" (I02), "evaluate" (I13), and "be transparent about mistakes" (I17) and believed that rigorous evaluation of collective intelligence was necessary to provide evidence of its usefulness to stakeholders, "so that it gets recognised and funded properly" (I13).

## Discussion

Our study showed that researchers were interested in looking for efficient methods of conducting research, leading them to try collective intelligence. Researchers believed that by involving large numbers of participants with various disciplines, they could find more innovative solutions to research problems in a shorter time with fewer costs compared to conventional methods. They indicated that participants' contributions could be solicited to solve problems, generate new research ideas, evaluate ideas, and create intellectual outputs. Researchers embarking on collective intelligence projects for the first time learned through the process and gradually improved their methods. They encountered barriers in planning and conducting collective intelligence projects due to the lack of a methodological guidance. We drew on the experiences of researchers across different fields and with experience of different collective intelligence methods to identify solutions and good practice advice to support researchers in the planning and implementation of their collective intelligence projects. This advice will help researchers prepare structures and processes for their projects, plan essential steps in their research, and foresee and develop strategies to overcome the barriers.

Despite increasing recognition of the value of collective intelligence in research [27,28], there are still examples of inappropriate methods used to mobilize collective intelligence [29]. For example, a project involving crowdsourcing in Rwanda failed to recruit and engage participants because the researchers mainly used social media for recruitment and requested participants to use a complicated tool for data collection [30]. However, community members in Rwanda were not connected on social media and were unfamiliar with the data collection tool. These issues could have been mitigated if the representatives of the target communities were involved from the outset as members of the project coordination team to advise on the conception and design of the collective intelligence project. A National Aeronautics and Space Administration competition to name a new node of the International Space Station was misled when an influential person called on the community to vote for his own name [31]. These examples emphasize the necessity of sharing experiences of researchers who have implemented collective intelligence projects to help future collective intelligence projects avoid methodological mistakes and outputs that are biased by group thinking.

Several efforts to define and standardize methods of collective intelligence in specific fields are available. These include a practical guide on using challenge contests to crowdsource ideas and solutions for health research from the World Health Organization and a list of toolkits compiled by the European Association of Citizen Science for researchers carrying out citizen science activities whereby members of the public collect and classify data [17,32]. However, a scoping review of the literature across different research fields classified four main methods to mobilize collective intelligence: independent contribution, challenge contest, games, and collaboration with a number of projects combining at least two methods [8]. By exploring experience of researchers who used one or more of these four methods in diverse disciplines, our study highlighted the barriers to mobilizing collective intelligence that researchers might encounter in different contexts. Good practice advice from researchers across disciplines could benefit researchers in planning and conducting future collective intelligence projects using one of these four methods within and outside health research.

To our knowledge, this is the first qualitative study to investigate the experiences of researchers in mobilizing collective intelligence across different fields. By using an online survey and semistructured interviews with a purposive sample of international researchers who had experience in implementing a range of different collective intelligence methods, we gained a breadth of perspectives. Respondents to the survey and interviews came from diverse disciplines, and some of them identified themselves as multidisciplinary researchers. The survey allowed a degree of interaction between researchers, which aided the analysis and interpretation of the results. Identification of areas that researchers agreed on helped us ascertain which barriers and strategies were applicable across different disciplines. Additionally, the semistructured interviews allowed researchers to explain the context of their research and describe their ideas and methods for addressing problems in mobilizing collective intelligence in depth.

Our study has some limitations. The online survey allowed participants to freely express their opinions, but we were unable to probe further to clarify the information written and gain a deeper understanding of their context. Furthermore, our survey and interview samples were mainly researchers who had published their collective intelligence projects. Therefore, we are uncertain about how far our findings are relevant to unpublished collective intelligence projects. Additionally, although we interviewed and surveyed researchers who had experience in running collective intelligence projects, we did not interview collective intelligence participants. Such data could provide further valuable insights on how to motivate and engage them.

In conclusion, mobilizing collective intelligence could be an effective way to improve research efficiency. The findings described in this paper should help researchers understand the barriers to implementing this new method. The good practice advice that we derived from respondents' accounts aims to support researchers in mobilizing collective intelligence effectively.

## Acknowledgments

This project is a part of a Methods in Research on Research (MiRoR) project that received funding from the European Union Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant (agreement no 676207). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Online survey questionnaire.

[PDF File (Adobe PDF File), 84KB - jmir_v21i7e13792_app1.pdf ]

## Multimedia Appendix 2

Interview guide.

[PDF File (Adobe PDF File), 77KB - jmir_v21i7e13792_app2.pdf ]

## Multimedia Appendix 3

Theme accumulation curve.

[PDF File (Adobe PDF File), 38KB - jmir_v21i7e13792_app3.pdf ]

## Multimedia Appendix 4

Advice that commentators disagreed with.

[PDF File (Adobe PDF File), 50KB - jmir_v21i7e13792_app4.pdf ]

**Multimedia Appendix 5**

Respondents' research disciplines.

[PDF File (Adobe PDF File), 70KB - jmir_v21i7e13792_app5.pdf ]

## References


1. Guinan E, Boudreau KJ, Lakhani KR. Experiments in Open Innovation at Harvard Medical School: What happens when an elite academic institution starts to rethink how research gets done? MIT Sloan Manag Rev 2013;54(3):45-52 [FREE Full text] [Medline: 29276450]
2. Zhang Y, Kim JA, Liu F, Tso LS, Tang W, Wei C, et al. Creative Contributory Contests to Spur Innovation in Sexual Health: 2 Cases and a Guide for Implementation. Sex Transm Dis 2015 Nov;42(11):625-628 [FREE Full text] [doi: 10.1097/OLQ.0000000000000349] [Medline: 26462186]
3. Malone T, Laubacher R, Dellarocas C. The collective intelligence genome. IEEE Eng Manag Rev 2010;38(3):38-52. [doi: 10.1109/EMR.2010.5559142]
4. Surowiecki J. The wisdom of crowds. New York: Anchor Books; 2005.
5. Bücheler T, Füchslin R, Pfeifer R. Crowdsourcing, Open Innovation and Collective Intelligence in the scientific method: a research agenda and operational framework. In: ALIFE. 2010 Presented at: Alife XII Conference; August 19, 2010; Odense, Denmark p. 679-686. [doi: 10.21256/zhaw-4094]
6. Howe J. Crowdsourcing: Why The Power Of The Crowd Is Driving The Future Of Business. New York: Crown Business; 2008.
7. Challenge.gov. About Challenge.gov URL: https://challenge.gov/a/buzz/pages/about-us [accessed 2019-01-23] [WebCite Cache ID 75dvhlb81]
8. Nguyen VT, Benchoufi M, Young B, Ghosn L, Ravaud P, Boutron I. A scoping review provided a framework for new ways of doing research through mobilizing collective intelligence. J Clin Epidemiol 2019 Jun;110:1-11 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.02.007] [Medline: 30772456]
9. Phylo. URL: https://phylo.cs.mcgill.ca/ [accessed 2019-01-15] [WebCite Cache ID 75RejICpE]
10. Foldit: Solve puzzles for science. URL: https://fold.it/portal/index.php?q= [accessed 2019-01-15] [WebCite Cache ID 75Rf2agMI]
11. Pianta MJ, Makrai E, Verspoor KM, Cohn TA, Downie LE. Crowdsourcing critical appraisal of research evidence (CrowdCARE) was found to be a valid approach to assessing clinical research quality. J Clin Epidemiol 2018 Dec;104:8-14 [FREE Full text] [doi: 10.1016/j.jclinepi.2018.07.015] [Medline: 30075189]
12. Küffner R, Zach N, Norel R, Hawe J, Schoenfeld D, Wang L, et al. Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. Nat Biotechnol 2015 Jan;33(1):51-57. [doi: 10.1038/nbt.3051] [Medline: 25362243]
13. Sims MH, Bigham J, Kautz H, Halterman MW. Crowdsourcing medical expertise in near real time. J Hosp Med 2014 Jul;9(7):451-456. [doi: 10.1002/jhm.2204] [Medline: 24740747]
14. Guinney J, Wang T, Laajala TD, Winner KK, Bare JC, Neto EC, Prostate Cancer Challenge DREAM Community. Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data. Lancet Oncol 2017 Dec;18(1):132-142 [FREE Full text] [doi: 10.1016/S1470-2045(16)30560-5] [Medline: 27864015]
15. Tang W, Han L, Best J, Zhang Y, Mollan K, Kim J, et al. Crowdsourcing HIV Test Promotion Videos: A Noninferiority Randomized Controlled Trial in China. Clin Infect Dis 2016 Jun 01;62(11):1436-1442. [doi: 10.1093/cid/ciw171] [Medline: 27129465]
16. Leiter A, Sablinski T, Diefenbach M, Foster M, Greenberg A, Holland J, et al. Use of crowdsourcing for cancer clinical trial development. J Natl Cancer Inst 2014 Oct;106(10):1-3. [doi: 10.1093/jnci/dju258] [Medline: 25217580]
17. Crowdsourcing in Health and Health Research: A Practical Guide. Geneva: World Health Organization; 2018:2018.
18. Créquit P, Mansouri G, Benchoufi M, Vivot A, Ravaud P. Mapping of Crowdsourcing in Health: Systematic Review. J Med Internet Res 2018 Dec 15;20(5):e187 [FREE Full text] [doi: 10.2196/jmir.9330] [Medline: 29764795]
19. European Citizen Science Association. URL: https://ecsa.citizen-science.net/about-us [accessed 2019-01-15] [WebCite Cache ID 75Rcy7rs3]
20. Collective intelligence 2018. URL: https://ci.acm.org/2018/program/ [accessed 2019-01-15] [WebCite Cache ID 75ReBwHcN]
21. Collective intelligence. URL: http://collectiveintelligenceconference.org/ [accessed 2019-01-15] [WebCite Cache ID 75ReWdcJQ]
22. Mailjet. URL: https://www.mailjet.com/ [accessed 2019-01-23] [WebCite Cache ID 75dcRPMr6]
23. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

25. Boeije H. A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. Quality and quantity 2002 Nov;36(4):391-409. [doi: 10.1023/A:1020909529486]
26. Tran V, Porcher R, Tran V, Ravaud P. Predicting data saturation in qualitative surveys with mathematical models from ecological research. J Clin Epidemiol 2017 Feb;82:71-78.e2. [doi: 10.1016/j.jclinepi.2016.10.001] [Medline: 27789316]
27. Mathews A, Farley S, Blumberg M, Knight K, Hightow-Weidman L, Muessig K, et al. HIV cure research community engagement in North Carolina: a mixed-methods evaluation of a crowdsourcing contest. J Virus Erad 2017 Oct 01;3(4):223-228 [FREE Full text] [Medline: 29057087]
28. Mathews A, Farley S, Hightow-Weidman L, Muessig K, Rennie S, Tucker J. Crowdsourcing and community engagement: a qualitative analysis of the 2BeatHIV contest. J Virus Erad 2018 Jan 01;4(1):30-36 [FREE Full text] [Medline: 29568551]
29. Tucker JD, Pan SW, Mathews A, Stein G, Bayus B, Rennie S. Ethical Concerns of and Risk Mitigation Strategies for Crowdsourcing Contests and Innovation Challenges: Scoping Review. J Med Internet Res 2018 Dec 09;20(3):e75 [FREE Full text] [doi: 10.2196/jmir.8226] [Medline: 29523500]
30. Jones M, Kondulis F. World Bank. 2016. Lessons from a crowdsourcing failure URL: https://blogs.worldbank.org/impactevaluations/lessons-crowdsourcing-failure [accessed 2019-01-14] [WebCite Cache ID 75QFjkTL3]
31. Zetlin M. Inc. 2016. The 5 worst crowdsourced ideas of all time URL: https://www.inc.com/minda-zetlin/the-5-biggest-crowdsourcing-fails-of-all-time-.html [accessed 2019-01-14] [WebCite Cache ID 75QHsbkUp]
32. European Citizen Science Association. 2019. A collection of Citizen Science guidelines and publications URL: https://ecsa.citizen-science.net/blog/collection-citizen-science-guidelines-and-publications [accessed 2019-05-07] [WebCite Cache ID 78BkODoEQ]



*Edited by G Eysenbach; submitted 25.02.19; peer-reviewed by S Day, J Tucker, W Tang; comments to author 11.04.19; revised version received 16.05.19; accepted 17.05.19; published 02.07.19.*

*Please cite as:*
*Nguyen VT, Young B, Ravaud P, Naidoo N, Benchoufi M, Boutron I*
*Overcoming Barriers to Mobilizing Collective Intelligence in Research: Qualitative Study of Researchers With Experience of Collective Intelligence*
*J Med Internet Res 2019;21(7):e13792*
*URL: https://www.jmir.org/2019/7/e13792/*
*doi:10.2196/13792*
*PMID:31267977*

Original Paper

# Computerized Quality of Life Assessment: A Randomized Experiment to Determine the Impact of Individualized Feedback on Assessment Experience

Daan Geerards[1,2,3], MD; Andrea Pusic[1,2], MD, MHS; Maarten Hoogbergen[3], MD, PhD; René van der Hulst[4], MD, PhD; Chris Sidey-Gibbons[1,2], PhD

[1]Patient-Reported Outcomes, Value & Experience Center, Brigham and Women's Hospital, Boston, MA, United States

[2]Department of Surgery, Harvard Medical School, Boston, MA, United States

[3]Department of Plastic and Reconstructive Surgery, Catharina Hospital, Eindhoven, Netherlands

[4]Department of Plastic and Reconstructive Surgery, Maastricht University Medical Center, Maastricht, Netherlands

**Corresponding Author:**
Chris Sidey-Gibbons, PhD
Patient-Reported Outcomes, Value & Experience Center
Brigham and Women's Hospital
75 Francis St
Boston, MA, 02115
United States
Phone: 1 6177328124
Email: drcgibbons@gmail.com

## Abstract

**Background:** Quality of life (QoL) assessments, or patient-reported outcome measures (PROMs), are becoming increasingly important in health care and have been associated with improved decision making, higher satisfaction, and better outcomes of care. Some physicians and patients may find questionnaires too burdensome; however, this issue could be addressed by making use of computerized adaptive testing (CAT). In addition, making the questionnaire more interesting, for example by providing graphical and contextualized feedback, may further improve the experience of the users. However, little is known about how shorter assessments and feedback impact user experience.

**Objective:** We conducted a controlled experiment to assess the impact of tailored multimodal feedback and CAT on user experience in QoL assessment using validated PROMs.

**Methods:** We recruited a representative sample from the general population in the United Kingdom using the Oxford Prolific academic Web panel. Participants completed either a CAT version of the World Health Organization Quality of Life assessment (WHOQOL-CAT) or the fixed-length WHOQOL-BREF, an abbreviated version of the WHOQOL-100. We randomly assigned participants to conditions in which they would receive no feedback, graphical feedback only, or graphical and adaptive text-based feedback. Participants rated the assessment in terms of perceived acceptability, engagement, clarity, and accuracy.

**Results:** We included 1386 participants in our analysis. Assessment experience was improved when graphical and tailored text-based feedback was provided along with PROMs (Δ=0.22, *P*<.001). Providing graphical feedback alone was weakly associated with improvement in overall experience (Δ=0.10, *P*=.006). Graphical and text-based feedback made the questionnaire more interesting, and users were more likely to report they would share the results with a physician or family member (Δ=0.17, *P*<.001, and Δ=0.17, *P*<.001, respectively). No difference was found in perceived accuracy of the graphical feedback scores of the WHOQOL-CAT and WHOQOL-BREF (Δ=0.06, *P*=.05). CAT (stopping rule [SE<0.45]) resulted in the administration of 25% fewer items than the fixed-length assessment, but it did not result in an improved user experience (*P*=.21).

**Conclusions:** Using tailored text-based feedback to contextualize numeric scores maximized the acceptability of electronic QoL assessment. Improving user experience may increase response rates and reduce attrition in research and clinical use of PROMs. In this study, CAT administration was associated with a modest decrease in assessment length but did not improve user experience. Patient-perceived accuracy of feedback was equivalent when comparing CAT with fixed-length assessment. Fixed-length forms are already generally acceptable to respondents; however, CAT might have an advantage over longer questionnaires that would be considered burdensome. Further research is warranted to explore the relationship between assessment length, feedback, and response burden in diverse populations.

*(J Med Internet Res 2019;21(7):e12212)* doi:10.2196/12212

**KEYWORDS**
quality of life; outcome assessment; patient-reported outcome measures; computer-adaptive testing; WHOQOL; psychometrics; feedback

## Introduction

### Background

Quality of life (QoL) assessments conducted using questionnaires are an important feature of clinical research and are increasingly being used to inform clinical practice. They have allowed psychologists, epidemiologists, and health care researchers to accurately quantify aspects relating to a person's QoL without relying on a structured interview with a trained professional. Though QoL questionnaires are commonly used in research studies and clinical trials, little research has been conducted to examine the effect of providing individualized feedback to people who complete these assessments, especially in QoL assessment [1,2].

In the context of health care provision, questionnaires that measure health and quality of life are often referred to as patient-reported outcome measures (PROMs). As an intervention designed to improve communication between patients and providers, PROMs can help health care providers understand what patients think about their own health. Gaining insight into patients' own appraisal of their health is important, as research demonstrates that clinicians may have limited insight into the effects of illness on patients' lives and cannot accurately predict how patients will rate their own mental and physical health [3,4]. PROMs are highly valued for their ability to address these problems [5-7].

Collection and feedback of PROMs in clinical practice can improve communication, decision making, satisfaction, and outcomes of care [8-11]. This information can be collected in many ways, ranging from basic paper-and-pen questionnaires to advanced computer systems. Research evidence suggests that only well-designed PROM interventions are likely to yield substantial improvements in clinical outcomes [12,13].

There are known barriers to using PROM questionnaires in both research and clinical practice. Doctors may avoid collecting PROMs because it can be difficult to relate to clinical decision making, and they fear it will add to clinical burden [5]. As patients or research participants, people may find questionnaires too burdensome or simply not interesting enough to justify completion [14].

### Burden

The burden of completing PROMs could be reduced by shortening assessments. Arbitrarily reducing the length of PROMs, however, would decrease the accuracy of the score estimates and the results. Computerized adaptive testing (CAT) is a technique that uses an algorithm to tailor questionnaire administration to individual patients and, as a result, is able to create short assessments while preserving accuracy [15,16].

Many simulation studies conducted in silico support the notion that CAT creates shorter assessments without sacrificing accuracy, and the assumption commonly presented is that shorter questionnaires always reduce burden, we are unaware of research that has evaluated the impact of shorter assessments on patient experience.

Along with reducing the length of questionnaires, the user experience of PROMs could be improved by making the assessment process more interesting and relevant. Advances in the power and availability of new computational tools indicate that CAT assessments can be readily deployed online alongside tailored graphical and text-based feedback. Research suggests that effective systems for collecting PROMs in clinical practice can be designed to capture information efficiently and provide clear feedback that makes it clear what should happen next; however, little is known about how these new technologies could be best used to improve user experience in this context [9,13,17].

### Objectives

In this study, we explore the impact of automatically generated personalized feedback on the user experience of electronic QoL assessment. We explore the following hypotheses:

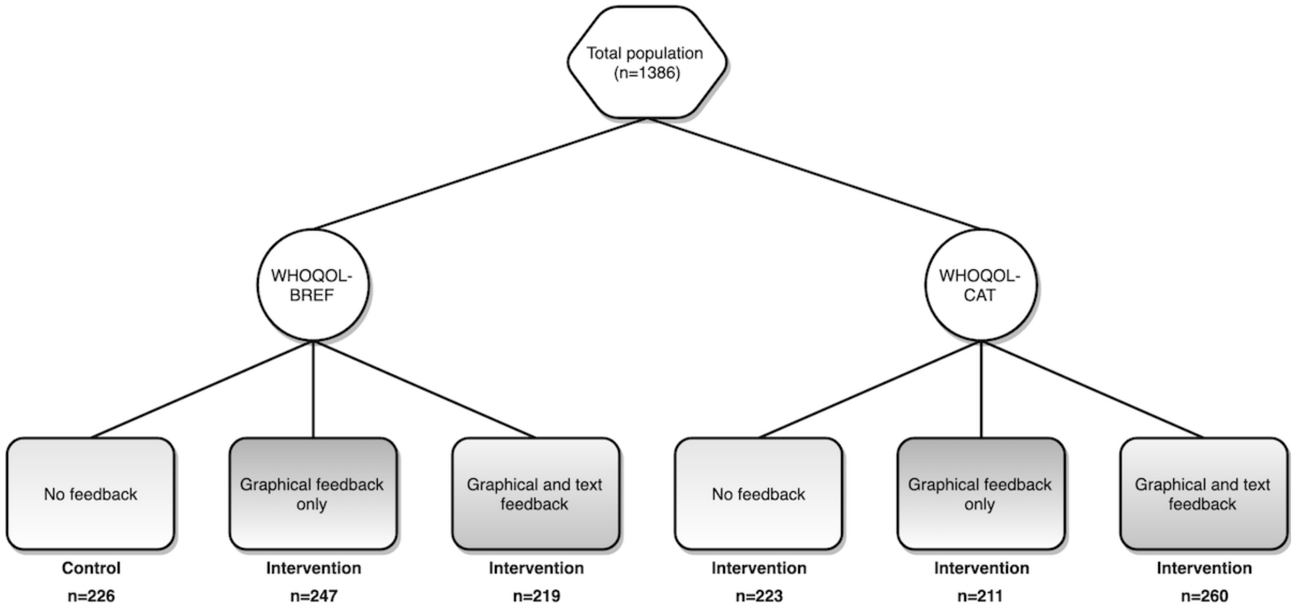
- Providing immediate feedback to the respondent will increase acceptability and satisfaction of the assessment.
- Contextualizing QoL scores using tailored text-based feedback will improve user experience compared with graphical feedback only.
- Perceived accuracy of graphical feedback scores is the same in CAT as in fixed-length assessment.
- CAT improves user experience and is shorter than the fixed-length questionnaires.

## Methods

### Study Sample

The sample consisted of participants from the general population in the United Kingdom. Participants were recruited between June 2017 and October 2017 through the Oxford Prolific Web panel, a crowdsourcing research platform [18]. We randomly assigned participants into 1 of 6 experimental conditions. In each condition, participants completed either a fixed-length QoL PROM or CAT QoL PROM [19-22]. In addition, they were randomly assigned to receive either no feedback, graphical feedback only, or graphical and adaptive tailored text-based feedback at the end of the questionnaire. Experimental conditions are displayed in Figure 1. Ethical approval was provided for this research by the institutional review board at the University of Cambridge Judge Business School (15-028).

**Figure 1.** Experimental conditions. WHOQOL-BREF: abbreviated version of World Health Organization Quality of Life-100; WHOQOL-CAT: computerized adaptive test version of World Health Organization Quality of Life-100.




## Measures

QoL assessments were based on the World Health Organization Quality of Life (WHOQOL) questionnaires [23-25]. The WHOQOL assesses different aspect of QoL, specifically physical health, psychological, social relationships, and environment. Responses are transformed to a score ranging from 0 (ie, worst QoL) to 100 (ie, best QoL). Each participant completed either a CAT version of the WHOQOL-100 (WHOQOL-CAT) or the fixed-length WHOQOL-BREF containing 24 items and 2 additional items assessing a respondent's overall perception of their QoL and health [19-22].

The WHOQOL-BREF was scored in accordance with published guidelines, the WHOQOL-CAT was scored using a maximum likelihood estimation of theta scores using a single parameter *Rasch* partial credit item response theory method with *Z*-score transformation based on norm scores from the UK population [19,22]. The stopping rule of the WHOQOL-CAT was set at an SE below 0.45.

After the QoL assessment, participants completed a survey to assess engagement and acceptability. The feedback samples completed additional items regarding feedback accuracy and clarity. Responses were scored on a 5-point Likert scale (ie, disagree strongly, disagree a little, neither agree or disagree, agree a little, and agree strongly). In addition, we collected data on time spent viewing feedback. All experimental stimuli were derived using the Web-based open-source Concerto software (University of Cambridge Psychometrics Centre) [26]. To control for guessing and cheating, the item "I have not been paying attention," was added, and respondents who endorsed this item were excluded from the analysis.

## Feedback

Graphical feedback was displayed as separate horizontal bar charts for each of the 4 WHOQOL domains, reflecting a score between 0 and 100. Text-based feedback included an explanation of what each domain reflects, how their score corresponds to average scores, and what their score might mean (eg, "Your score of 26 on this scale indicates that your psychological quality of life is lower than average. This suggests that your satisfaction with your psychological health is lower than it could be. You may be worrying more than usual, struggling to make decisions or not feeling content with your life. You could discuss these things with your doctor"). Feedback was augmented with a series of geographically relevant hyperlinks (assuming that the participant allowed their browser to access details of their location) to signpost relevant support services for each of the 4 domains. An example of how feedback was shown to participants can be seen in Multimedia Appendix 1.

## Analysis

All analyses were conducted within the R Statistical Programming Environment (version 3.4.4) [27]. Descriptive statistics were derived for age, gender, working status, and mean WHOQOL scores. We compared the research gold standard of a fixed-length QoL assessment without feedback (ie, control sample) with the other 5 conditions that are presented in Figure 1. The mean score and SD were derived for each survey item, including a summary score of all engagement and acceptability items (ie, total score of the 4 survey items). Separate item scores ranged from 0 to 4, where a score of 0 corresponded to the response "disagree strongly," and a score of 4 corresponded to the response "agree strongly" (ie, 5-point Likert scale). The overall assessment score had a range of 0 to 16. Effect size for ordinal data was derived with Cliff delta (mean and range). Significance was assessed by performing Wilcoxon tests with a cutoff of $P<.005$ to increase reproducibility [28-31]. As proposed by Benjamin et al, a score of $P<.05$ was defined as suggestive instead of significant [30]. Time spent looking at feedback was compared between the *only graphical feedback* and the *graphical & text-based feedback* conditions.

We used the *mokken* package in R to perform Mokken scale analysis on the 4 acceptability and engagement items to assess unidimensionality and scalability. Scalability was displayed as

Loevinger coefficient H, where a scale is considered weak if H<.3 and strong if H>.5. Unidimensionality was assessed by finding potential Mokken scales, with a cutoff set at .3 [32,33]. Furthermore, internal consistency of all 4 items were reflected with Cronbach alpha, derived by using the *psych* package in R [34]. A Cronbach alpha >.70 is generally seen as satisfactory when comparing groups [35,36].

# Results

## Study Sample

In total, 1454 participants completed the questionnaire. After excluding 68 respondents who endorsed the item "I have not been paying attention," 1386 respondents were remaining for the analysis. Descriptive statistics (age, gender, working status, and mean WHOQOL scores) are presented in Table 1. The population distribution for all 6 conditions can be seen in Figure 1.

**Table 1.** Demographics (n=1386).

| Characteristics | Statistics |
|---|---|
| **Age (years)** | |
| Mean (SD) | 40 (12) |
| Range | 18-75 |
| **Gender, n (%)** | |
| Female | 669 (48.3) |
| Male | 544 (39.2) |
| Not reported | 173 (12.5) |
| **Working status, n (%)** | |
| Full-time paid work (≥30 hours/week) | 556 (40.1) |
| Part-time paid work (<30 hours/week) | 226 (16.3) |
| Full-time education at school, college, or university | 64 (4) |
| Looking after home | 134 (9) |
| Fully retired from work | 77 (5) |
| Permanently sick or disabled | 58 (4) |
| Unemployed | 55 (4) |
| **Mean World Health Organization Quality of Life scores, mean (SD)** | |
| Physical | 68 (18) |
| Psychological | 59 (17) |
| Social | 57 (17) |
| Environmental | 72 (13) |

Scale and item properties Mokken analysis showed that all items were loaded on a single component, meaning that all items were assessing acceptability. Furthermore, Loevinger H was found to be .53 for the total scale, with all items having values >.30. Cronbach alpha was found to be .77.

## Overall Feedback and User Experience

When combining the WHOQOL-CAT and WHOQOL-BREF samples, providing graphical and tailored text-based feedback significantly improved overall experience compared with no feedback ($mean_{no\ feedback}$ 11.2, SD 3.0; $mean_{graphical\ and\ text}$ 12.3, SD 2.7; $P \leq .001$; Δ=0.22, Δ 95% CI 0.15-0.28). Providing only graphical feedback had a suggestive effect on overall user experience ($mean_{no\ feedback}$ 11.2, SD 3.0; $mean_{graphical\ feedback}$ 11.8, SD 2.9; $P$=.006; Δ=0.10, Δ 95% CI 0.02-0.17).

Text-based and graphical feedback was also found to improve user experience when comparing all samples separately. Furthermore, respondents thought the questionnaire with graphical and text-based feedback was more interesting compared with no feedback assessment, whereas providing only graphical feedback did not make the questionnaire more interesting. Participants who received graphical and text-based feedback were also more likely to report they would share the questionnaire with someone else. All results are presented in Table 2, in which every separate sample was compared with the control sample (fixed-length assessment without feedback).

**Table 2.** Assessment survey results. All samples are compared with the World Health Organization Quality of Life-BREF no feedback control sample.

| Variable | World Health Organization Quality of Life-BREF | | | Computerized adaptive test version of World Health Organization Quality of Life-100 | | |
|---|---|---|---|---|---|---|
| | No feedback (n=226) | Graphical feedback (n=247) | Graphical & text-based feedback (n=219) | No feedback (n=223) | Graphical feedback (n=211) | Graphical & text-based feedback (n=260) |
| **Overall assessment (total of 4 items)** | | | | | | |
| Score (SD) | 11.5 (2.88) | 11.98 (2.74) | *12.31 (2.54)* [b] | 10.99 (3.09) | 11.57 (3.13) | *12.35 (2.89)* |
| Wilcoxon *P* | —[a] | .07 | *.002* | .08 | .72 | *<.001* |
| Δ[c] (range) | — | 0.09 (−0.01 to 0.19) | *0.16 (0.06 to 0.26)* | 0.11 (−0.21 to 0.00) | 0.01 (−0.09 to 0.12) | *0.18 (0.09 to 0.27)* |
| **"The questionnaire was interesting"** | | | | | | |
| Score (SD) | 3.22 (0.79) | 3.26 (0.81) | *3.43 (0.75)* | 3.02 (0.92) | 3.30 (0.81) | *3.46 (0.70)* |
| Wilcoxon *P* | — | .41 | *<.001* | .03 | .17 | *<.001* |
| Δ (range) | — | 0.04 (−0.05 to 0.13) | *0.17 (0.07 to 0.26)* | −0.11 (−0.21 to −0.01) | 0.07 (−0.03 to 0.17) | *0.17 (0.08 to 0.26)* |
| **"I am satisfied with the amount of information"** | | | | | | |
| Score (SD) | 3.29 (0.84) | 3.35 (0.73) | 3.44 (0.69) | 3.19 (0.81) | 3.30 (0.82) | 3.37 (0.83) |
| Wilcoxon *P* | — | .66 | .07 | .11 | .99 | .20 |
| Δ (range) | — | 0.02 (−0.08 to 0.11) | 0.08 (−0.01 to 0.18) | −0.08 (−0.18 to 0.01) | −0.01 (−0.10 to 0.09) | 0.06 (−0.03 to 0.15) |
| **"It would be useful to share with someone else; perhaps my friends, spouse, or doctor"** | | | | | | |
| Score (SD) | 2.12 (1.19) | 2.43 (1.06) | *2.48 (1.04)* | 2.11 (1.17) | 2.24 (1.21) | *2.51 (1.08)* |
| Wilcoxon *P* | — | .006 | *.001* | .92 | .32 | *<.001* |
| Δ (range) | — | 0.14 (0.04 to 0.24) | *0.17 (0.07 to 0.27)* | −0.01 (−0.11 to 0.10) | 0.05 (−0.05 to 0.16) | *0.18 (0.09 to 0.28)* |
| **"I would recommend this questionnaire to a friend"** | | | | | | |
| Score (SD) | 2.90 (0.93) | 2.94 (0.96) | 2.97 (0.92) | 2.69 (1.05) | 2.76 (1.11) | 2.98 (1.02) |
| Wilcoxon *P* | — | .53 | .38 | .03 | .35 | .16 |
| Δ (range) | — | 0.03 (-0.07 to 0.12) | 0.04 (−0.06 to 0.14) | −0.12 (−0.22 to −0.02) | −0.05 (−0.15 to 0.05) | 0.07 (−0.03 to 0.16) |

[a]Contains no results since this was the control group for comparison with the other samples.

[b]Italicized results are significant (*P*<.005).

[c]Δ=Cliff delta.

No difference was found in perceived accuracy of the graphical feedback scores of the WHOQOL-CAT and WHOQOL-BREF ($\text{mean}_{\text{CAT feedback accuracy}}$ 2.9, SD 0.9; $\text{mean}_{\text{fixed feedback accuracy}}$ 3.1, SD 1.0; *P*=.05; Δ=0.06, Δ 95% CI 0.00-0.12). Furthermore, 757 out of 919 (82.4%) participants thought the graphical feedback was accurate, and 850 out of 915 (92.9%) participants thought the graphical feedback was clear. In the text-based feedback sample, 384 out of 469 (81.9%) participants affirmed accuracy of text-based feedback and 445 out of 468 (95.1%) affirmed clearness of text-based feedback. Response distribution of feedback appraisal is shown in Table 3.

**Table 3.** Feedback accuracy and clarity responses.

| Feedback response | Responses, n | Disagree, n (%) | Neutral, n (%) | Agree, n (%) |
|---|---|---|---|---|
| The graphical feedback was accurate | 919 | 83 (9.0%) | 79 (8.6%) | 757 (82.4%) |
| The graphical feedback was clear | 915 | 27 (2.9%) | 38 (4.2%) | 850 (92.9%) |
| The text feedback was accurate | 469 | 50 (10.7%) | 35 (7.5%) | 384 (81.9%) |
| The text feedback was clear | 468 | 7 (1.5%) | 16 (3.4%) | 445 (95.1%) |

### Computerized Adaptive Testing

CAT did not improve overall assessment experience scores compared with the fixed-length ($mean_{CAT}$ 11.7, SD 3.1; $mean_{fixed}$ 11.9, SD 2.7; *P*=.21; Δ=−0.06, Δ 95% CI −0.12 to 0.00). Even when combining adaptive assessment with graphical feedback, assessment experience did not significantly differ from the fixed-length assessment without feedback ($mean_{CAT\_graphical}$ 11.6, SD 3.1; $mean_{fixed\_nofeedback}$ 11.5, SD 2.9; *P*=.72; Δ=0.01, Δ 95% CI −0.09 to 0.12).

In the WHOQOL-CAT sample, mean items administered was 17.9 (SD 2.3), compared with 24 items in the WHOQOL-BREF, which corresponds to an item reduction of 25.4%.

### Feedback Time

Median time spent looking at feedback for all feedback groups combined was 129 seconds. Respondents in the WHOQOL-CAT sample spent significantly more time looking at graphical and text-based feedback compared with graphical feedback only, which is shown in Table 4. In the WHOQOL-BREF group, the difference in time looking at feedback did not comply to our *P* value threshold but has suggestive significance.

**Table 4.** Time spent looking at feedback. All samples are compared with the World Health Organization Quality of Life-BREF graphical feedback control sample.

| Variable | World Health Organization Quality of Life-BREF | | Computerized adaptive test version of World Health Organization Quality of Life-100 | |
|---|---|---|---|---|
| | Graphical feedback (n=247) | Graphical & text-based feedback (n=219) | Graphical feedback (n=211) | Graphical & text-based feedback (n=260) |
| **Time spent looking at feedback** | | | | |
| Median, seconds | 115 | 132 | 124 | *147* [a] |
| Wilcoxon *P* | __[b] | .016 | .42147 | *<.001* |
| Δ[c] (range) | — | 0.13 (0.02 to 0.23) | 0.04 (−0.06 to 0.15) | *0.24 (0.15 to 0.34)* |

[a]Italicized results are significant (*P*<.005).

[b]Contain no results since this was the control group for comparison with the other samples.

[c]Δ=Cliff delta.

## Discussion

### Conclusions

With this study, we have shown that immediately providing feedback after online QoL assessment significantly improves assessment experience when providing combined graphical and tailored text-based feedback. Graphical feedback alone did not improve assessment experience. Perceived accuracy of feedback was not different when comparing WHOQOL-BREF with WHOQOL-CAT, which suggests that CAT scores are as reliable as fixed-length scores, from a respondent's perspective. The WHOQOL-CAT is shorter than fixed-length assessment, but it did not necessarily result in a better experience. Furthermore, respondents thought both graphical and text-based feedback after WHOQOL assessment were considerably clear and accurate.

### Other Literature

Little research has been conducted to assess QoL assessment feedback. Brundage et al assessed interpretation accuracy and ratings of ease of understanding and usefulness of different data presentation formats for both patients and respondents in both group-level data and individual-level data [1]. They looked at how graphical data should be provided and with which details, where we looked at, and what kind of feedback, including tailored text-based feedback, is most desired and found to be accurate by respondents. Kuijpers et al assessed self-rated understanding of Quality of Life Questionnaire-Core 30 scores and preference for presentation styles, whereas our research did not focus on presentation style but on feedback method [2].

### Strengths

This study has several strengths. The sample size is relatively large, with sufficient distribution in age, gender, and working status to accurately reflect the British population. The sample was large enough to establish 5 different large samples for comparison with our control group. By establishing 6 different samples, we were better able to target our different comparisons and hypotheses. We accepted a probability value of *P*=.005, rather than the conventional *P*=.05, to increase the likelihood of these results being replicable in future investigations. As the use of a *P* value of .005 might not be widely accepted yet, we have regarded a conventional *P* value of .05 as suggestive instead of significant [28-31].

We took steps to ensure our data were of high quality by adding dummy questions to the questionnaire to assess attentiveness and removing participants who stated that they were not paying attention, in line with recommendations for increasing reliability in studies conducted using compensated Web panels [37].

### Acceptability and Engagement Evaluation

In this study, we focused on the impact of feedback and impact of CAT on the user experience and assessed perceived accuracy of graphical feedback. As this represents 1 of the first efforts to do so, we were unable to find a previously validated questionnaire for assessing the relevance and acceptability of feedback and CAT administration. We developed a questionnaire that used items that had been shown to work well in the single other study we found that examined acceptability of feedback for patients completing a personality questionnaire [38]. Though we found that the questionnaire performed well during psychometric evaluation, we acknowledge that the short questionnaire may not cover all relevant aspects of questionnaire completion.

### Computerized Adaptive Testing

The questionnaire was, primarily, designed to assess experience relating to feedback (eg, the questionnaire was interesting; it would be useful to share with someone else, perhaps a friend, spouse, or doctor), which partially explains the lack of effect between the CAT and the fixed-length groups. In this study, the average for WHOQOL-CAT (SE<0.45) items administered was 17.9, which is only slightly more than expected based on an earlier simulation study conducted in silico (average 16.7, SE <0.45) [22]. This moderate item reduction in an already brief questionnaire might also explain why CAT did not affect user experience in this study. We know fixed-length questionnaires are likely to be acceptable to respondents, proven by their long-standing effective use. However, CAT is likely to provide an advantage over fixed-length forms that would be considered burdensome. Further research is therefore warranted to explore the relationship between CAT, questionnaire length, and patient burden.

### Assessment and Acceptability

Some unanticipated results were found. The scores on each questionnaire item were positively skewed in each group, indicating that although feedback did significantly improve experience, there appears to be something inherently positive about completing the WHOQOL questionnaire, regardless of the provision of feedback. Our experimental design prohibits us from understanding if the recruitment method (ie, via Web panel with compensation) affected the scores of acceptability measure, though participants were aware their reimbursement was not linked in any way to their responses. Studies that have been designed to assess the reliability of responses from similar Web panels (eg, Amazon's Mechanical Turk) have found them to be reliable sources of information though, but to our knowledge, no studies have focused specifically at either QoL research or participants from the United Kingdom [39]. For evaluation of the assessment, we created our own survey items and conducted psychometric analyses to assess their suitability. Alternative for assessing feedback from clinical assessments are available, for example the "Patient Feedback Form" developed by Basch et al and adapted by Snyder et al, but unfortunately, this form has not been psychometrically validated for use in the English language. After we finished our data inclusion, Tolstrup et al translated and validated this evaluation form for a Danish patient population. Future research in this area may usefully validate the Patient Feedback Form for use in English [40-42].

### Impact

We discovered that the effect sizes were small to moderate. Despite the modesty of the effect sizes, we consider that adding tailored text-based feedback to outcome assessment might have a considerable impact on user experience, engagement, and response rate when feedback is implemented in outcome assessments where the primary goal is to maximize response rates and minimize longitudinal attrition. We chose a cross-sectional design to investigate this effect, and our positive results suggest that further experimentation in a cohort of patients who are prospectively followed up using QoL PROMs is warranted. In this study, we compared a participant's scores to the population mean; during longitudinal assessment it becomes possible to feedback a person's scores in relation to their previous scores, which may further increase the relevance of individualized feedback and therefore, the acceptability and willingness to participate in multiple PROM assessments over time.

### Bottom Line

In conclusion, providing feedback after outcome assessments is important to maximize user experience. Putting scores into context by using tailored text increased the user engagement. In addition, in this study, CAT did not improve overall experience but was substantially shorter than the fixed-length assessment. More research is necessary to assess CAT patient burden in terms of PROM assessment time, item reduction, patient-perceived length, and patient-perceived validity.

## Acknowledgments

Funding was received for this study from the National Institute for Health Research (CDF 2017-10-19).

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Example of feedback.

[PDF File (Adobe PDF File), 7MB - jmir_v21i7e12212_app1.pdf ]

## References

1. Brundage MD, Smith KC, Little EA, Bantug ET, Snyder CF, PRO Data Presentation Stakeholder Advisory Board. Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. Qual Life Res 2015 Oct;24(10):2457-2472 [FREE Full text] [doi: 10.1007/s11136-015-0974-y] [Medline: 26012839]

2. Kuijpers W, Giesinger JM, Zabernigg A, Young T, Friend E, Tomaszewska IM, et al. Patients' and health professionals' understanding of and preferences for graphical presentation styles for individual-level EORTC QLQ-C30 scores. Qual Life Res 2016 Mar;25(3):595-604 [FREE Full text] [doi: 10.1007/s11136-015-1107-3] [Medline: 26353905]
3. Devlin NJ, Appleby J, Buxton M. Getting the Most Out of PROMs : Putting Health Outcomes at the Heart of NHS Decision-Making. London, UK: King's Fund; 2010.
4. Kwoh CK, O'Connor GT, Regan-Smith MG, Olmstead EM, Brown LA, Burnett JB, et al. Concordance between clinician and patient assessment of physical and mental health status. J Rheumatol 1992 Jul;19(7):1031-1037. [Medline: 1512756]
5. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. Br Med J 2015 Feb 10;350:g7818. [doi: 10.1136/bmj.g7818] [Medline: 25670183]
6. Dowrick C, Leydon GM, McBride A, Howe A, Burgess H, Clarke P, et al. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. Br Med J 2009 Mar 19;338:b663. [doi: 10.1136/bmj.b663] [Medline: 19299474]
7. Black N. Patient reported outcome measures could help transform healthcare. Br Med J 2013 Jan 28;346:f167. [doi: 10.1136/bmj.f167] [Medline: 23358487]
8. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 2013 Jun 11;13:211 [FREE Full text] [doi: 10.1186/1472-6963-13-211] [Medline: 23758898]
9. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008 Mar;17(2):179-193. [doi: 10.1007/s11136-007-9295-0] [Medline: 18175207]
10. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006 Oct;12(5):559-568. [doi: 10.1111/j.1365-2753.2006.00650.x] [Medline: 16987118]
11. Santana MJ, Feeny D, Johnson JA, McAlister FA, Kim D, Weinkauf J, et al. Assessing the use of health-related quality of life measures in the routine clinical care of lung-transplant patients. Qual Life Res 2010 Apr;19(3):371-379. [doi: 10.1007/s11136-010-9599-3] [Medline: 20146009]
12. Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, et al. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database Syst Rev 2016 Jul 13;7:CD011119 [FREE Full text] [doi: 10.1002/14651858.CD011119.pub2] [Medline: 27409972]
13. Krägeloh CU, Czuba KJ, Billington DR, Kersten P, Siegert RJ. Using feedback from patient-reported outcome measures in mental health services: a scoping study and typology. Psychiatr Serv 2015 Mar 01;66(3):224-241. [doi: 10.1176/appi.ps.201400141] [Medline: 25727110]
14. Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R, et al. Increasing response rates to postal questionnaires: systematic review. Br Med J 2002 May 18;324(7347):1183 [FREE Full text] [doi: 10.1136/bmj.324.7347.1183] [Medline: 12016181]
15. Gershon RC. Computer adaptive testing. J Appl Meas 2005;6(1):109-127. [Medline: 15701948]
16. Wainer H, Dorans NJ, Flaugher R, Green BF, Mislevy RJ. Computerized Adaptive Testing: A Primer. London, UK: Routledge; 2000.
17. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? Qual Life Res 2009 Feb;18(1):115-123. [doi: 10.1007/s11136-008-9430-6] [Medline: 19105048]
18. Prolific. URL: https://prolific.ac/ [accessed 2018-09-10] [WebCite Cache ID 72Kd2avAl]
19. The WHOQOL group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998 May;28(3):551-558. [Medline: 9626712]
20. Skevington SM, Lotfy M, O'Connell KA, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004 Mar;13(2):299-310. [Medline: 15085902]
21. Berlim MT, Pavanello DP, Caldieraro MA, Fleck MP. Reliability and validity of the WHOQOL BREF in a sample of Brazilian outpatients with major depression. Qual Life Res 2005 Mar;14(2):561-564. [Medline: 15892446]
22. Gibbons C, Bower P, Lovell K, Valderas J, Skevington S. Electronic quality of life assessment using computer-adaptive testing. J Med Internet Res 2016 Sep 30;18(9):e240 [FREE Full text] [doi: 10.2196/jmir.6053] [Medline: 27694100]
23. WHOQOL Group. Development of the WHOQOL: rationale and current status. Int J Ment Health 1994;23(3):24-56. [doi: 10.1080/00207411.1994.11449286]
24. The WHOQOL group. The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med 1995 Nov;41(10):1403-1409. [Medline: 8560308]
25. World Health Organization. The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 1998 Jun;46(12):1569-1585. [Medline: 9672396]
26. Concerto. URL: http://concertoplatform.com/ [accessed 2018-09-10] [WebCite Cache ID 72KbAOfke]
27. R Project. The R Project for Statistical Computing URL: https://www.r-project.org/ [accessed 2018-09-10] [WebCite Cache ID 72KbUInJV]

28. Ioannidis JP. The proposal to lower P value thresholds to .005. J Am Med Assoc 2018;319(14):1429-1430. [doi: 10.1001/jama.2018.1536] [Medline: 29566133]
29. Colquhoun D. The reproducibility of research and the misinterpretation of p-values. R Soc Open Sci 2017 Dec;4(12):171085 [FREE Full text] [doi: 10.1098/rsos.171085] [Medline: 29308247]
30. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. Nat Hum Behav 2018 Jan;2(1):6-10. [doi: 10.1038/s41562-017-0189-z] [Medline: 30980045]
31. Johnson VE. Revised standards for statistical evidence. Proc Natl Acad Sci U S A 2013 Nov 26;110(48):19313-19317. [doi: 10.1073/pnas.1313476110] [Medline: 24218581]
32. Mokken RJ. A Theory and Procedure of Scale Analysis. Berlin, Germany: Walter de Gruyter; 1971.
33. van der Ark LA. New developments in Mokken scale analysis in R. J Stat Softw 2012;48(5):1-27. [doi: 10.18637/jss.v048.i05]
34. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, PROMIS Cooperative Group. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). Med Care 2007 May;45(5 Suppl 1):S22-S31. [doi: 10.1097/01.mlr.0000250483.85507.04] [Medline: 17443115]
35. Bland JM, Altman DG. Cronbach's alpha. Br Med J 1997 Feb 22;314(7080):572 [FREE Full text] [doi: 10.1136/bmj.314.7080.572] [Medline: 9055718]
36. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
37. Rouse SV. A reliability analysis of mechanical Turk data. Comput Hum Behav 2015 Feb;43:304-307. [doi: 10.1016/j.chb.2014.11.004]
38. Lengel GJ, Mullins-Sweatt SN. The importance and acceptability of general and maladaptive personality trait computerized assessment feedback. Psychol Assess 2017;29(1):1-12. [doi: 10.1037/pas0000321] [Medline: 27099976]
39. Goodman JK, Cryder CE, Cheema A. Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. J Behav Decis Mak 2012;26(3):213-224. [doi: 10.1002/bdm.1753]
40. Tolstrup LK, Pappot H, Zangger G, Bastholt L, Zwisler AD, Dieperink KB. Danish translation, cultural adaption and initial psychometric evaluation of the patient feedback form. Health Qual Life Outcomes 2018 Apr 27;16(1):77 [FREE Full text] [doi: 10.1186/s12955-018-0900-4] [Medline: 29703191]
41. Snyder CF, Blackford AL, Wolff AC, Carducci MA, Herman JM, Wu AW, PatientViewpoint Scientific Advisory Board. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901 [FREE Full text] [doi: 10.1002/pon.3087] [Medline: 22544513]
42. Basch E, Artz D, Dulko D, Scher K, Sabbatini P, Hensley M, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. J Clin Oncol 2005 May 20;23(15):3552-3561 [FREE Full text] [doi: 10.1200/JCO.2005.04.275] [Medline: 15908666]


## Abbreviations

**CAT:** computerized adaptive testing
**PROMs:** patient-reported outcome measures
**WHOQOL:** World Health Organization Quality of Life
**WHOQOL-BREF:** abbreviated version of WHOQOL-100
**WHOQOL-CAT:** computerized adaptive test version of World Health Organization Quality of Life-100

*Edited by G Eysenbach; submitted 14.09.18; peer-reviewed by C Snyder, J Alvarez Pitti; comments to author 22.11.18; revised version received 28.03.19; accepted 02.05.19; published 11.07.19.*

*Please cite as:*
*Geerards D, Pusic A, Hoogbergen M, van der Hulst R, Sidey-Gibbons C*
*Computerized Quality of Life Assessment: A Randomized Experiment to Determine the Impact of Individualized Feedback on Assessment Experience*
*J Med Internet Res 2019;21(7):e12212*
*URL: http://www.jmir.org/2019/7/e12212/*
*doi:10.2196/12212*
*PMID:31298217*

Viewpoint

# Artificial Intelligence and the Implementation Challenge

James Shaw[1,2], PhD; Frank Rudzicz[3], PhD; Trevor Jamieson[1,4], MD, MBI; Avi Goldfarb[5], PhD

[1]Women's College Hospital, Institute for Health System Solutions and Virtual Care, Toronto, ON, Canada
[2]Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada
[3]International Centre for Surgical Safety, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada
[4]St Michael's Hospital, Toronto, ON, Canada
[5]Rotman School of Management, University of Toronto, Toronto, ON, Canada

**Corresponding Author:**
James Shaw, PhD
Women's College Hospital
Institute for Health System Solutions and Virtual Care
76 Grenville Street
Toronto, ON, M5G2A2
Canada
Phone: 1 4163236400
Email: jay.shaw@wchospital.ca

## Abstract

**Background:** Applications of artificial intelligence (AI) in health care have garnered much attention in recent years, but the implementation issues posed by AI have not been substantially addressed.

**Objective:** In this paper, we have focused on machine learning (ML) as a form of AI and have provided a framework for thinking about use cases of ML in health care. We have structured our discussion of challenges in the implementation of ML in comparison with other technologies using the framework of Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies (NASSS).

**Methods:** After providing an overview of AI technology, we describe use cases of ML as falling into the categories of decision support and automation. We suggest these use cases apply to clinical, operational, and epidemiological tasks and that the primary function of ML in health care in the near term will be decision support. We then outline unique implementation issues posed by ML initiatives in the categories addressed by the NASSS framework, specifically including meaningful decision support, explainability, privacy, consent, algorithmic bias, security, scalability, the role of corporations, and the changing nature of health care work.

**Results:** Ultimately, we suggest that the future of ML in health care remains positive but uncertain, as support from patients, the public, and a wide range of health care stakeholders is necessary to enable its meaningful implementation.

**Conclusions:** If the implementation science community is to facilitate the adoption of ML in ways that stand to generate widespread benefits, the issues raised in this paper will require substantial attention in the coming years.

***(J Med Internet Res 2019;21(7):e13659)*** doi:10.2196/13659

**KEYWORDS**

artificial intelligence; machine learning; implementation science; ethics

## Introduction

Artificial intelligence (AI) has become a topic of central importance to the ways in which health care will change in the coming decades, with recent commentaries addressing potential transformations in clinical care [1,2], public health [3], and health system planning [4]. AI is a general purpose technology (GPT), which means it represents a core set of capabilities that can be leveraged to perform a wide variety of tasks in different contexts of application [5]. Understanding the core capabilities of AI as a GPT, and the ways in which it stands to be incorporated into health care processes, is essential for the implementation research community to contribute to promoting a positive place for AI in the future of health care. We believe that AI has the potential to substantially reconfigure health care, with implications that reach beyond enhancing the efficiency and effectiveness of care delivery. Due to this potential, we suggest that implementation science researchers and practitioners make a commitment to more fully consider the wider range of issues that relate to its implementation, which include health system, social, and economic implications of the deployment of AI in health care settings.

We suggest that the most appropriate language for discussions of AI in health care is actually to discuss machine learning (ML), which is the specific subfield of AI that is currently making the most impact across industries. We then focus on 2 questions about the deployment of ML in health care. First, how should ML be understood in terms of its actual use cases in health care? This question addresses the nature of ML as an *implementation object* [6,7] in health-related contexts. We present a basic framework for thinking about use cases of ML in terms of decision support versus automation and elaborate clinical, operational, and epidemiological categories of these use cases.

Second, what are the unique challenges posed by ML that may require consideration during an implementation initiative? As opposed to focusing on strategies for the adoption of digital technologies in general, which has been addressed extensively in other literature [8-10], we focus on what we understand to be the most important risks arising from the implementation of ML in health care. Our discussion of the risks associated with implementing ML in health care is guided by the work of Greenhalgh et al in the framework for theorizing and evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability (NASSS) of health and care technologies [8].

The NASSS framework is based on the premise that when considering influences on whether and how a technology is successfully taken up and used, it is important to keep in mind that "it is not individual factors that make or break a technology implementation effort but the dynamic interaction between them" [8]. The NASSS framework outlines a range of considerations that are relevant to understanding how a technology might be adopted across an entire region or health system, ranging from a focus on the particular health condition in the clinical scenario to the wider political, regulatory, and sociocultural system in which it is to be embedded. In our paper, we examine ML as a GPT that has the potential to apply across clinical conditions and focus our analysis on elements of the NASSS framework: the technology, its value propositions, and the adopters, organizations, and systems into which it might be introduced. We emphasize the evolutionary nature of ML as a GPT and explicitly acknowledge that it will continue to develop and change over the coming years, which is also an important feature of the NASSS framework. We conclude by advocating for further research on the risks posed by ML from an implementation science perspective.

AI has been described in many ways. Using the framing in Agrawal et al, we emphasize that recent advances in AI can be best understood as "prediction technology" [11]. Quite simply, prediction is defined for this purpose as "taking information you have, often called 'data', and using it to generate information you don't have" (PM, p. 24). This newly generated information *estimates* the true information that is missing, leading to the potential for people and technology to take actions that may have otherwise been based on less accurate information.

Predicting illness episodes that might be experienced in the future is an obvious application of AI in this sense, but prediction as we have defined it has many other uses as well. Examples include an automatic translator predicting the phrases of Spanish that correspond to a particular set of phrases in English or a chat bot predicting the most appropriate cluster of words in response to a given query. These examples might not represent the very intuitive understanding of prediction that we have become used to in everyday usage or the way we tend to think of prediction of health-related events and outcomes in health care. However, they represent the prediction of information that we do not have based on information we do have and point toward the potentially widespread applications of AI as a GPT.

The phrase "predictive analytics" is very intuitive with regard to defining AI as a prediction technology, using advanced computer algorithms to predict health-related events from existing data in ways that exceed the ability of individual researchers applying individual analyses [12]. However, AI opens new opportunities for prediction beyond the familiar predictive analytics for hospital admissions, length of stay, and patient survival rates. As a process of filling in missing information, better and cheaper prediction is already being used in new areas, from transcribing audio to enhancing security to informing diagnoses.

At its core, current applications of AI bring statistical modeling, computer code, and advanced computing power to bear on large amounts of representative data. In his recent commentary on the potential of deep learning (a form of AI) to transform health care, Hinton gave the example of deciding whether a patient has a particular disease and explained that a common approach would be to use a simple logistic regression (using data to predict a binary outcome: the patient has the disease or does not). However, he suggested that if there are extremely high numbers of potential influences or predictors of whether the person has the disease, many of which may interact with one another, the prediction challenge becomes much more complex. This is especially the case where we have imperfect knowledge of the causes and correlates of a particular disease. This example also pertains only to binary queries specifically about whether a patient has a single disease, which is different from the typical reasoning processes involved in differential diagnosis among clinicians, where multiple confounding, interdependent outcomes must be considered [13,14].

Specific applications of AI can fall under distinct categories, with AI serving as an *umbrella* concept, covering more specific frameworks. In this paper, we are primarily concerned with the subdomain of AI referred to as ML in which statistical models are automatically (or semiautomatically) induced from data according to some criterion (eg, best expected discriminative power or maximum likelihood given to training data). This means that complex statistical models capable of executing advanced predictions are generated in part by using data to *train* the model to achieve a particular goal.

Often, ML involves *supervised* methods that categorize data points, for example, as images of skin cancer or otherwise given datasets in which all data points (or at least a substantial subset) are associated with a label, ordinal, or category that is meant to be predicted or inferred [15]. This process requires datasets that have the appropriate labels indicating what the data means; in the example of images of skin cancer, each data point would be labeled according to its representation of a mass as *malignant* or *benign* or some variation thereof. Given these labels and the statistical models they help to train, ML can be very effective at determining the category in which any newly available individual data point belongs, thereby being useful in the effort to, for example, identify malignant cancers based on particular images [16].

Much of the power of modern ML also derives from *unsupervised* pattern recognition, in which hidden (or *latent*) aspects of the data are automatically identified by the algorithms and exploited according to the aforementioned criteria. Unsupervised ML can often identify patterns in the data that humans do not even think to look for. Often, these hidden aspects are nonlinear combinations of many parts of the input.

ML can also improve its ability to *take actions* according to these induced hidden patterns and particular functions of cost and reward in a process called *reinforcement learning*. For example, ML can dynamically adapt survey questions to more quickly identify possible diseases [17], dynamically avoid potential communication breakdowns during speech conversation in the assessment of dementia [18], and even recommend treatments directly when using structured institutional data [19]. As so much health care information can be represented digitally, the potential of ML to improve health care practices is profound.

## Methods

### Use Cases of Machine Learning in Health Care

In the remainder of our paper we refer primarily to ML as opposed to AI, focusing our analysis on the concrete possibilities of ML in health care. We can think about use cases of ML in health care in 2 broad ways. The first is through *decision support*, wherein ML algorithms are used to provide some form of input into human decision making. An example is where an algorithm is used to provide more accurate predictions of the outcome of a particular procedure given a particular clinical presentation. This helps to inform a human decision about whether a given procedure is the best course of action. The second is through *automation*, wherein algorithms are used not only to predict an output but also to take action to achieve a particular outcome. An example is the automatic transcribing of a clinical note when dictated into a computer program, resulting in a complete note being added to a patient's record (technically referred to as Automated Speech Recognition). These 2 broadly defined categories of use cases can be thought of as applying to various types of tasks in health care, and we suggest it is instructive to consider 3 types of tasks as most relevant for the implementation of ML for health: clinical, operational, and epidemiological.

Clinical tasks refer to health-related assessment, intervention, and evaluation, generally performed by qualified health care providers, for example, determining a differential diagnosis. Operational tasks are those related to activities that are ancillary to clinical tasks but necessary or valuable in the delivery of services, such as generating, storing, and retrieving medical records. Finally, epidemiological tasks are those related to more accurately identifying the health needs and outcomes of a set of people within a given population. An example is the development of a warning system for disease outbreak. As epidemiological use cases of ML are related to enhancing the ability of humans to make decisions in the other categories described here (clinical or operational), there are no examples of pure automation for epidemiological tasks that contain an output other than informing a human decision. Hypothetical examples of both decision support and automation are given under each of these categories in Table 1.

This table presents a basic framework for thinking about use cases of ML in health care as falling into 2 primary categories: decision support and automation. These use cases apply in categories of clinical, operational, and epidemiological tasks. As no examples of pure automation exist for epidemiological tasks, no example is presented in that cell.

The considerations most pertinent to the implementation of ML will depend on the particular use case being proposed in a given implementation initiative, and the categories outlined in Table 1 provide a framework for understanding those use cases. The NASSS framework and other work in implementation science for digital health technologies emphasize the importance of attending to the particular value proposition that a new technology offers for health care stakeholders [8,9]. The value proposition of digital technology might be different for different stakeholder groups, and implementation frameworks direct attention to the implications of newly introduced technologies for patients, health care providers, managers, health policymakers, and others [8,25,26]. The clinical, operational, and epidemiological task types presented in Table 1 will correspond to different value propositions for different stakeholder groups, meaning that specific applications of ML might preferentially benefit one group over another, for example, identifying a scheduling process to maximize efficiency in operating costs might preferentially benefit managers over health care providers inconvenienced by a new system. Understanding how value propositions differ for the various stakeholders implicated in a given implementation of ML is an essential consideration for successful adoption and use.

**Table 1.** Examples of use cases in each category of application.

| Type of use case | Clinical[a] | Operational[b] | Epidemiological[c] |
|---|---|---|---|
| Decision support | Producing a more accurate prediction of the likely outcome of a particular intervention [20] | Identifying potential staff scheduling changes related to forecasted emergency room volumes [21] | Warning systems for disease outbreak [22] |
| Automation | Automatically altering insulin treatment in response to monitored glucose-insulin dynamics [23] | Use of robotics for operational tasks in dementia care, such as meal delivery [24] | N/A[d] |

[a]Tasks related to the assessment, intervention, and evaluation of health-related issues and procedures, generally performed by qualified health care providers.

[b]Tasks related to activities that are ancillary to clinical tasks but necessary or valuable in the delivery of services.

[c]Tasks related to more accurately identifying the health needs and outcomes of people within a given population.

[d]Not applicable.

The potential value propositions of an ML technology offering decision support versus one offering automation are very different and bring along different sorts of implementation issues. The implementation of decision support systems in health care that do not include applications of ML have been well studied and the difficulties include perceived challenges to autonomy, lack of time, and dissatisfaction with user interfaces [27,28]. Implementation initiatives involving decision support applications of ML will need to consider this past work to develop implementation strategies that more effectively address known challenges.

Implementation initiatives involving automation are likely to face some similar and some different challenges. For example, stakeholder views on the introduction of automated robotics into a variety of health care settings found a widespread lack of interest and understanding and fear of the ways in which work would be disrupted and distributed [29]. Although automation has existed in health care for decades through technologies such as heart rate monitors, the question of how acceptable stakeholders will perceive new forms of automation to be remains an important issue. This point raises the overarching issue of the extent of automation that is possible through applications of ML, linked to speculation about whether ML will mostly *augment* or actually *replace* health care providers' work [1,30].

### Augmentation and Replacement of Health Care Work

We agree with a growing chorus of health care providers and researchers who suggest that ML will primarily serve to *augment* as opposed to *replace* the work of humans in the provision of health care in the near term [31], despite applications of automation in health care. This is because the role of ML in the current generation of capabilities functions at the level of the *task*, and not at the level of an entire *job*. Agrawal et al explained that "the actual implementation of AI is through the development of tools. The unit of AI tool design is not 'the job' or 'the occupation' or the 'the strategy', but rather 'the task'." (p. 125). Therefore, for a health care provider to be entirely replaced, every single task performed by that provider would need to be automated by an ML tool or handed off to a different human.

The complete automation of the full range of human tasks involved in providing clinical care is not yet possible; activities such as making treatment decisions based on a differential diagnosis that integrates data from laboratory investigation, visual observation, and patient history are still too complex for automation. In emphasizing this point, we are suggesting that although much of the hype about AI (and specifically ML) in health care has focused on its potential role in *automating* processes of health service delivery, it is more likely that near-term applications of ML will fall under the category of *decision support*.

Further comments about prediction tasks and decision tasks will help to clarify this point. As stated earlier, ML applications fundamentally perform some form of prediction. The specific instance of prediction that the application is performing may be thought of as the *prediction task*, which may be paired with a complementary *decision task*. The decision task is where the newly generated information is used to select a particular action in a given context. In applications of ML that function as decision support, the decision task is performed by a human. As ML diffuses, an important new challenge for health care providers is to make choices using the predictions that arise from decision support applications of ML, involving new forms of input to clinical thought processes related to risks, benefits, and previously unrecognized influences on health. The examples of decision support in Table 1 involve generating better information to inform human decision making.

In applications of ML that function as automation, both the prediction task and the decision task are accomplished by machines. A clear example is self-driving cars. The sensors surrounding the car enable predictions of the best direction in which the car should travel. However, it is the *selection* from a predetermined set of actions and *execution* of one action over another that makes self-driving cars an example of automation as distinct from one of decision support. ML is not yet sophisticated enough to complete these selection and execution functions for many health care tasks, across both clinical and operational levels.

As prediction tasks become more amenable to being performed by ML, decision tasks become more valuable [5,32]. This is because predictions are improved, meaning that decisions can be made with greater confidence and impact. The enhanced value of these decisions represents the potential value of ML as a decision support tool and illustrates the potential breadth of value propositions that could arise from this technology with a wide range of implications for the implementation process.

However, for decision support to be valuable in health care, the outputs of algorithms must have a clear entryway into the human decision-making processes that pervade health service delivery. This points us toward one of a series of important issues raised from an implementation science perspective on the introduction of ML in health care settings, which we turn to next.

## Results

### Unique Considerations for Implementation Science

We have described use cases (and attendant value propositions) of ML in health care as more likely relating to decision support and less likely to automation, which begins to illustrate the implementation object of focus in ML initiatives [6,7]. In many cases of decision support, the implementation object is actually not all that different from the statistical tools that are already used as part of common practice, such as risk prediction. In cases of automation, there are similarly many examples of technologies that have already been successfully implemented in health care settings (such as automatic transcription mentioned earlier). However, ML as a GPT raises a number of issues that run across use cases and might be anticipated as unique in comparison with implementation projects for other digital technologies.

Best practices of implementation for digital innovations [8,9,33] will be fundamental to the adoption of ML in health care. Here, we discuss considerations that might appear in implementation projects involving ML that may be less likely to appear in implementation projects involving other digital technologies and yet stand to have a potentially strong influence on the success of such projects. We organize this section based on distinct levels of consideration that are presented in the NASSS framework that we have not yet addressed [8,26]: health care providers, patients and the public, health care organizations, and health policy and systems. Although we consider the primary considerations of health technology vendors working on the development of ML application in health care to be outside the scope of this paper, we acknowledge this is a gap in the literature that requires attention.

### Health Care Providers

Health care providers are those responsible for doing the actual work of health care delivery and are being increasingly expected to adopt and use new technologies in health care environments. We suggest that the core considerations or risks of the implementation of ML for health care providers will fall into the categories of meaningful decision support and explainability.

#### *Meaningful Decision Support*

For ML to function as decision support in a way that is valuable to health care stakeholders, the outputs of algorithms must have a meaningful entryway into decision making. From an operational or epidemiological perspective, isolated analyses of risk prediction may help to inform resource allocation and subsequent analysis decisions fairly simply. However, from a clinical perspective, algorithms that perform isolated risk prediction may be less useful. Clinical decision making is a complex process involving the integration of a variety of data sources, incorporating both tacit and explicit modes of intelligence [34-36]. To inform this decision-making process more intuitively, attention is increasingly being devoted to communication tools such as data visualization [37]. The nature and value of these communication tools are central to the implementation process, helping to determine whether and how algorithmic outputs are incorporated in everyday routine practices. This point primarily relates to the decision support use case across clinical, operational, and epidemiological tasks.

#### *Explainability*

There is a growing concern in the AI community related to the explainability of the results achieved by ML algorithms, wherein the ways in which algorithms enhance the performance of prediction can often not be understood [38]. As a result of the processes described earlier in this paper, the ways in which data are being used to train algorithms cannot be traced out in sequential, logical detail. Hence, the actual ways in which models achieve their results are in some instances not knowable even to the computer scientists who create them. Evidence-based medicine rests on a foundation of the highest standards of explainability; medical decision making aspires to incorporate a sound understanding of the mechanisms by which diseases and their treatments function and the particular treatments that have demonstrated the greatest benefits under particular experimental circumstances (in addition to patient needs and values [35,39,40]). The lack of understanding of those mechanisms and circumstances poses challenges to the acceptability of ML to health care stakeholders. Although the issue of explainability relates clearly to decision support uses cases of ML as explained here, the issue may apply even more profoundly to automation-focused use cases as they gain prominence in health care.

### Patients and the Public

The issues of public trust and public input into the governance of ML initiatives in health care have been widely discussed as the popularity of AI has grown, with advocates suggesting that future developments of AI ought to be explicitly supporting a broader public interest. We suggest that 2 pairs of issues frame the risks of ML related to patients and the public. The first pair is privacy and consent and the second is representative data and algorithmic bias.

#### *Privacy and Consent*

The training of ML models requires large amounts of data, which means that applications of ML in health will likely rely on health-related data from patients and the public. As governments and other actors internationally become interested in developing applications of ML, health-related data are increasingly made available to private entities with the capability of producing AI applications that are relevant to peoples' health [41-43]. Currently, data from wearable devices such as smart watches and mobile apps are not widely covered by health information legislation [44], and many health-related apps have unclear consenting processes related to the flow of data generated through their use [45]. Furthermore, data that are de-identified may be reidentifiable when linked with other datasets [46]. These considerations create major risks for initiatives that seek to make health data available for use in the development of ML applications, potentially leading to substantial resistance from health care providers such as that seen in primary care in Denmark in recent years [42]. This will be particularly important for population and public health use cases that require data from very large segments of the population. The meaning of consent and strategies to maintain patient privacy are central considerations to ML implementation initiatives. The related issues of privacy and consent pertain especially to clinical and epidemiological use cases of ML in both decision support and automation categories, as data from patients /or the public are essential to train algorithms in these areas (whereas operational use cases may only rely on other forms of data, such as clinical scheduling histories).

### *Representative Data and Algorithmic Bias*

Algorithms are only as good as the data used to train them. In cases where training data are partial or incomplete or only reflect a subset of a given population, the resulting model will only be relevant to the population of people represented in the dataset [47]. This raises the question about data provenance [30,48] and represents a set of issues related to the biases that are built into algorithms used to inform decision making. One high profile example was the hiring bias exhibited when algorithms were used to make hiring decisions at Amazon, resulting in only men being advanced to subsequent stages of hiring [49]. This is notable in part because the algorithm performed extremely well based on the available data, simply extending the bias that already existed in hiring practices at the company. When applied to health care of public health, data provenance and potential bias in training data represent important issues that are likely to be of major concern for the stakeholders involved in the implementation of an ML initiative. Public health has health equity as a primary goal, and representativeness in terms of which populations can be addressed by an ML initiative will be a central consideration.

A further challenge with the nature of the data on which algorithms are trained relates to *concept drift*, a phenomenon where data on which an algorithm is trained change over time (or become out of date), which changes the performance of the algorithm as new data are acquired [50]. The possibility of concept drift means that those overseeing the performance of ML-based technologies in health care must identify strategies to determine how well the algorithm deals with new data and whether concept drift is occurring. Applications to support this effort are emerging in the literature [51].

The issues addressed here apply most clearly to ML applications that use patient data to inform clinical and epidemiological use cases that enhance clinical care and health system planning. And although the use of public data will likely be the most contentious issue in this domain, the challenges of representativeness and bias apply to all ML use cases across decision support and automation domains.

## Health Care Organizations

Health care and public health systems are composed of independent organizations that need to develop and execute strategies within the limits of the resources available to them. Organizations have been the driving force behind the adoption of many innovations in health care and have a collection of considerations that are unique from the broader systems of which they are a part. We suggest that the issue of security and computational resources become particularly important for organizations as they adopt ML initiatives in health care and public health.

### *Security*

As data are collated and stored for training ML models, the risk and potential severity of security breaches grows. The global attack of health care organizations using *WannaCry* ransomware in May 2017 shows the vulnerabilities of even well protected health data to malicious interests. This particular attack is estimated to have affected 200,000 systems in over 150 countries, indicating the potential scope of security problems as the value of data grows [52]. Strategies to prevent such security breaches on Web accessible health data are now being proposed in the literature [53,54], and the high profile of security issues makes this a particularly important issue as ML applications develop in health care and public health. The issue of security transcends any particular use case of ML and includes any applications or analysis that relies on big data more generally.

### *Computational Resources*

Advanced applications of ML require substantial computing power, with some predictive analyses and training models requiring up to several weeks to run. The more extensive the computing support, the more efficient ML applications will become, raising the question of the cost and availability of such advanced computing power for health care organizations. Health care is publicly funded in many countries around the world, and public support to secure the resources to fund the necessary computing power may not be present. Cloud-based analytics present an opportunity and a challenge for health-related organizations in relation to the issue of computational resources. Cloud-based data analysis means that organizations would not need to own computational resources directly [55] but also introduces the potential challenges of data safety. These issues are relevant to the training phase of a newly developed algorithm, but of course, less computing power is required to simply apply algorithms that have been generated and trained elsewhere. How data are stored and processed is thus also an important consideration in ML implementation initiatives. The issue of computational resources also applies more generally than any given ML use case, related to the development and functioning of many kinds of AI algorithms.

## Health Policy and Systems

The challenges associated with ML initiatives at the level of health policy and systems are extensive. These include broad legislative frameworks related to emerging health-related technologies more generally [56] and to the innovation procurement systems that vary across health system settings [57,58]. The policy issues presented by ML in health care are beginning to garner more attention [42,43], but here we present one issue that we have not seen addressed in health care or public health literature: the challenge of scalability.

### *Scalability and Normal Accidents*

A major challenge that extends beyond any single implementation of ML, and therefore requires a system-wide view, relates to the *scalability* of ML. Scalability in this sense refers to the unanticipated effects of the appearance of multiple ML technologies that will inevitably interact with one another by some means. As applications of ML proliferate across health care and public health, eventually some algorithmic outputs will confront others. The effects of this interaction are impossible to predict in advance, in part because the particular technologies that will interact are unclear and likely not yet implemented in the course of usual care.

Health care represents what Charles Perrow referred to as a complex system or a system in which processes are tightly linked to one another and interact in unintended ways in the effort to achieve the goals of the system [59]. This acknowledgement has led to the high reliability movement in health care and other industries [60], intending to implement management strategies that could mitigate against the risk of disasters arising from such immense complexity. Perrow's work was titled *Normal Accidents: Living with High Risk Technologies*, suggesting that in systems characterized by complexity and the use of advanced technologies, accidents are bound to happen [59]. This basic point about the seeming inevitability of accidents in the context of complex systems and new technologies underscores the significance of the scalability challenge of ML in health care. We suggest that implementation scientists will need to consider the unintended consequences of the implementation and scale of ML in health care, creating even more complexity and greater opportunity for risks to the safety of patients, health care providers, and the general public. ML safety will likely need to become a dedicated focus of patient safety research internationally. This point about scalability frames the broader challenge for implementation scientists who are committed to a system-wide perspective on health innovations and relates not only to each type of use case identified in our framework but also to the interactions between them as well.

# Discussion

## Intersecting Issues in the Future of Health Care

In our brief Discussion section, we outline 2 overarching issues that we consider to frame the challenges facing health care systems that are hoping to adopt ML in the coming years. The discussion here is informed by the explicit recognition in the NASSS framework that both the technology and context in which innovations are being introduced shift and change over time. Greenhalgh et al suggest that although the levels of the framework can be distinguished analytically, "at an empirical level they are inextricably interlinked and dynamically evolving, often against a rapidly shifting policy context or continued evolution of the technology" (p. 14). Our assessment of the 2 issues we address here is intended to represent the connections between the changes that will be required as the policy context and technology evolve concurrently. The first is the issue of the role of corporations in health-related applications of ML, and the second is the issue of the role of ML in the evolving nature of health care.

## The Role of Corporations

As the innovations enabled by ML have taken on a more powerful role in driving global economies, corporations have strategically sought to acquire larger amounts of more diverse data to boost their capacity to develop ML algorithms [61]. The shifting focus of many large corporations to the collection and manipulation of data characterizes what Zuboff refers to as *surveillance capitalism*, a relatively recent phenomenon in the global economy that relies on data for innovation and corporate success. The more that large corporations enter the health care industry with the power to collect, store, and use data, the more intertwined health care will become with the corporate realities of these large, multinational companies [62].

As large corporations acquire more data and develop more sophisticated forms of ML that transcend any individual geographical region, the implications for domestic health care policy are at risk of being overlooked. Although recent efforts to create regional protections around data collection and use have appeared to make an impact, such as the General Data Protection Regulation in Europe, health care policy is well behind. In cases where health-related data are already being stored in a country other than where the user is living, what are the regulations on how those data can be used? Where users voluntarily engage with technologies that collect their data for explicit health-related use by a corporation outside of their political jurisdiction, what legislative frameworks apply to protect patients and the public? These issues represent the important challenge of making health policy matter when conventional political boundaries are less able to contain the potential of large corporations to develop and use their technological capabilities.

## The Changing Nature of Artificial Intelligence–Enabled Health Care

AI applications represent a potential impetus for major change in the institutions that constitute health care. In this sense, the term institution refers not just to the organizations in which health care providers work but to a complex collection of cognitive, cultural, regulative, and moral influences that shape the way that health care workers see their work and their lives [63]. The social sciences have worked to provide clear definitions of institutions through decades of research and theory [63-65]. Scott explained that institutions are combinations of 3 pillars: norms of *the way things are usually done around here* (cultural-cognitive influences), laws and regulations (regulative influences), and assumed moral codes (normative influences) [63]. Health care represents a confluence of institutions understood in this sense, many of which are naturally oriented toward maintaining some version of the status quo. Particularly for members of institutions who maintain power over resources, such as the medical profession, embracing institutional change is a point of resistance and difficulty.

We suggest that ML will confront the realities of entrenched institutions through issues such as meaningful decision support and explainability described earlier. These 2 issues represent the authority of health care providers over the decisions that come to define health care as a multi-institutional field, both in terms of their rightful positions within the system and the fabric of decision making that has always defined health care processes. These issues point toward an important challenge that we suggest implementation scientists must grapple with: the changing nature of health care work. In *Prediction Machines*, the authors explain that as AI technology develops, "the value of substitutes to prediction machines, namely human prediction, will decline. However, the value of complements, such as the human skills associated with data collection, judgment, and actions, will become more valuable." (p. 81). As the implementation science community considers how to encourage the adoption of ML technologies, it will also need to consider how such technologies stand to change the ways in which health care planning, decision making, and delivery are understood and the evolving role of human health care providers within that context.

The challenges described here refer to unique considerations of ML that pose novel challenges to implementation beyond the work of promoting the routine use of technologies among health care providers. We suggest that the hype and high stakes of ML make these issues more prominent in the mindsets of health care stakeholders and therefore more likely to impact upon an ML implementation project. The implementation science community will need to establish strategies to address these issues as ML becomes more prominent, each of which requires ongoing work to be adequately addressed.

## Conclusions

In this paper, we have provided an overview of ML for implementation scientists informed by the NASSS framework, outlining the use cases of ML as falling into the categories of decision support and automation. We suggest these use cases apply to clinical, operational, and epidemiological tasks and that the primary ways in which ML will enter into health care in the near term will be through decision support. We then outlined unique implementation issues posed by ML initiatives from 4 perspectives, those of health care providers, patients and the public, health care organizations, and health policy and systems.

Ultimately, we suggest that the future of ML in health care remains positive but uncertain, as support from patients, the public, and a wide range of health care stakeholders is necessary to enable its meaningful implementation. However, as applications of ML become more sophisticated and investment in communications strategies such as data visualization grows, ML is likely to become more user-friendly and more effective. If the implementation science community is to facilitate the adoption of ML in ways that stand to benefit all, the issues raised in this paper will require substantial attention in the coming years.

## Acknowledgments

This research is supported by an Associated Medical Services Phoenix Fellowship to the corresponding author.

## Authors' Contributions

JS led the writing of the manuscript. JS, TJ, AG, and FR contributed to the conceptualization, design, and approach for the manuscript. JS, TJ, AG, and FR contributed to analysis and interpretation of the argument made in the manuscript. All authors contributed to writing and revising the manuscript. All authors provided the final approval of the manuscript. All authors agree to be accountable for the manuscript.

## Conflicts of Interest

None declared.

## References

1. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. J Am Med Assoc 2016 Dec 13;316(22):2353-2354. [doi: 10.1001/jama.2016.17438] [Medline: 27898975]
2. Naylor CD. On the prospects for a (deep) learning health care system. J Am Med Assoc 2018 Sep 18;320(11):1099-1100. [doi: 10.1001/jama.2018.11103] [Medline: 30178068]
3. Thiébaut R, Thiessard F, Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics. Artificial intelligence in public health and epidemiology. Yearb Med Inform 2018 Aug;27(1):207-210. [doi: 10.1055/s-0038-1667082] [Medline: 30157525]
4. Harwich E, Laycock K. Wilton Park. London: National Health Service; 2018. Thinking On Its Own: AI in The NHS URL: https://www.wiltonpark.org.uk/wp-content/uploads/Thinking-on-its-own-AI-in-the-NHS.pdf
5. Ajay A, Gans J, Goldfarb A. The Economics of Artificial Intelligence: An Agenda. Washington DC: National Bureau of Economic Research; 2019.
6. Nilsen P, Ståhl C, Roback K, Cairney P. Never the twain shall meet?-a comparison of implementation science and policy implementation research. Implement Sci 2013;8(1):63. [doi: 10.1186/1748-5908-8-63]
7. 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]

8. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Dec 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
9. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond 'implementation': digital health innovation and service design. NPJ Digit Med 2018;1(1):48. [doi: 10.1038/s41746-018-0059-8]
10. May C. Agency and implementation: understanding the embedding of healthcare innovations in practice. Soc Sci Med 2013;78:26-33. [doi: 10.1016/j.socscimed.2012.11.021published] [Medline: 23246396]
11. Agrawal A, Gans J, Goldfarb A. Prediction Machines: The Simple Economics of Artificial Intelligence. Brighton, Massachusetts: Harvard Business Review Press; 2018.
12. Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff (Millwood) 2014 Jul;33(7):1139-1147. [doi: 10.1377/hlthaff.2014.0048] [Medline: 25006139]
13. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ 2005 Jan;39(1):98-106. [doi: 10.1111/j.1365-2929.2004.01972.x] [Medline: 15612906]
14. Victor-Chmil J. Critical thinking versus clinical reasoning versus clinical judgment: differential diagnosis. Nurse Educ 2013;38(1):34-36. [doi: 10.1097/NNE.0b013e318276dfbe] [Medline: 23222632]
15. Esteva A, Kuprel B, Novoa R, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
16. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17 [FREE Full text] [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
17. Rajpurkar P, Polamreddi V, Balakrishnan A. ResearchGate.: Cornell University Library; 2017. Malaria likelihood prediction by effectively surveying households using deep reinforcement learning URL: https://www.researchgate.net/publication/321324506_Malaria_Likelihood_Prediction_By_Effectively_Surveying_Households_Using_Deep_Reinforcement_Learning
18. Chinaei H, Currie LC, Danks A, Lin H, Mehta T, Rudzicz F. Identifying and avoiding confusion in dialogue with people with alzheimer's disease. Comput Linguist 2017 Jun;43(2):377-406. [doi: 10.1162/COLI_a_00290]
19. Liu N, Logan B, Liu N, Xu Z, Tang J, Wang Y. Deep reinforcement learning for dynamic treatment regimes on medical registry data. Healthc Inform 2017 Aug;2017:380-385 [FREE Full text] [doi: 10.1109/ICHI.2017.45] [Medline: 29556119]
20. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 2016 Mar;3(3):243-250. [doi: 10.1016/S2215-0366(15)00471-X] [Medline: 26803397]
21. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. Acad Emerg Med 2008 Feb;15(2):159-170 [FREE Full text] [doi: 10.1111/j.1553-2712.2007.00032.x] [Medline: 18275446]
22. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. IEEE Access 2017;5:8869-8879. [doi: 10.1109/ACCESS.2017.2694446]
23. Miller S, Nimri R, Atlas E, Grunberg EA, Phillip M. Automatic learning algorithm for the MD-logic artificial pancreas system. Diabetes Technol Ther 2011 Oct;13(10):983-990. [doi: 10.1089/dia.2010.0216] [Medline: 21774690]
24. Casey D, Beyan O, Murphy K, Felzmann H. Robot-Assisted Care for Elderly With Dementia: Is There a Potential for Genuine End-User Empowerment? In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts. 2015 Presented at: HRI'15 Extended Abstracts; March 2-5, 2015; Portland, OR p. 247-248.
25. Chaudoir SR, Dugan AG, Barr CH. Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. Implement Sci 2013 Feb 17;8:22 [FREE Full text] [doi: 10.1186/1748-5908-8-22] [Medline: 23414420]
26. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
27. Légaré F, O'Connor AM, Graham ID, Saucier D, Côté L, Blais J, et al. Primary health care professionals' views on barriers and facilitators to the implementation of the Ottawa decision support framework in practice. Patient Educ Couns 2006 Nov;63(3):380-390. [doi: 10.1016/j.pec.2006.04.011] [Medline: 17010555]
28. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. J Am Med Inform Assoc 2017 Mar 1;24(2):423-431. [doi: 10.1093/jamia/ocw105] [Medline: 27516495]
29. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res 2018;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]
30. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med 2018;15(11):e1002689 [FREE Full text] [doi: 10.1371/journal.pmed.1002689] [Medline: 30399149]
31. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]

32. Agrawal A, Gans J, Goldfarb A. WSB Wiki. Toronto: University of Toronto; 2018. The Labor Market Impact of Artificial Intelligence URL: https://wiki.bus.wisc.edu/download/attachments/45908081/AGG_2018_10_15.pdf?version=1&modificationDate=1549310002037&api=v2
33. Shaw J, Shaw S, Wherton J, Hughes G, Greenhalgh T. Studying scale-up and spread as social practice: theoretical introduction and empirical case study. J Med Internet Res 2017;19(7):e244 [FREE Full text] [doi: 10.2196/jmir.7482] [Medline: 28687532]
34. Greenhalgh T, Wieringa S. Is it time to drop the 'knowledge translation' metaphor? A critical literature review. J R Soc Med 2011 Dec;104(12):501-509 [FREE Full text] [doi: 10.1258/jrsm.2011.110285] [Medline: 22179293]
35. Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? Br Med J 2014 Jun 13;348:g3725 [FREE Full text] [doi: 10.1136/bmj.g3725] [Medline: 24927763]
36. Patel VL, Kaufman DR, Arocha JF. Emerging paradigms of cognition in medical decision-making. J Biomed Inform 2002 Feb;35(1):52-75 [FREE Full text] [doi: 10.1016/S1532-0464(02)00009-6] [Medline: 12415726]
37. Rossi RA, Ahmed NK. The Network Data Repository with Interactive Graph Analytics and Visualization. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015 Presented at: AAAI'15; January 25-30, 2015; Austin Texas p. 4292-4293.
38. Samek W, Wiegand T, Müller KR. arXiv. 2017. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models URL: https://arxiv.org/pdf/1708.08296.pdf
39. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. J Am Med Assoc 1992 Nov 4;268(17):2420-2425. [doi: 10.1001/jama.1992.03490170092032] [Medline: 1404801]
40. Tracy CS, Dantas GC, Upshur RE. Evidence-based medicine in primary care: qualitative study of family physicians. BMC Fam Pract 2003 May 9;4:6 [FREE Full text] [doi: 10.1186/1471-2296-4-6] [Medline: 12740025]
41. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 2017;7(4):351-367 [FREE Full text] [doi: 10.1007/s12553-017-0179-1] [Medline: 29308344]
42. Wadmann S, Hoeyer K. Dangers of the digital fit: rethinking seamlessness and social sustainability in data-intensive healthcare. Big Data Soc 2018 Jan 11;5:205395171775296. [doi: 10.1177/2053951717752964]
43. Vezyridis P, Timmons S. Understanding the care.data conundrum: new information flows for economic growth. Big Data Soc 2017 Jan;4(1):205395171668849. [doi: 10.1177/2053951716688490]
44. Gostin LO, Halabi SF, Wilson K. Health data and privacy in the digital era. J Am Med Assoc 2018 Jul 17;320(3):233-234. [doi: 10.1001/jama.2018.8374] [Medline: 29926092]
45. Grundy Q, Held FP, Bero LA. Tracing the potential flow of consumer data: a network analysis of prominent health and fitness apps. J Med Internet Res 2017;19(6):e233 [FREE Full text] [doi: 10.2196/jmir.7347] [Medline: 28659254]
46. Culnane C, Rubinstein B, Teague V. arXiv. 2017. Health data in an open world URL: https://arxiv.org/ftp/arxiv/papers/1712/1712.05627.pdf
47. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. Big Data Soc 2016 Dec;3(2):205395171667967. [doi: 10.1177/2053951716679679]
48. Yu KH, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019 Mar;28(3):238-241. [doi: 10.1136/bmjqs-2018-008551] [Medline: 30291179]
49. Higginbottom K. Forbes. 2018. The pros and cons of algorithms in recruitment URL: https://www.forbes.com/sites/karenhigginbottom/2018/10/19/the-pros-and-cons-of-algorithms-in-recruitment/#79f3658b7340
50. Tsymbal A. ResearchGate.: Computer Science Department, Trinity College Dublin; 2004. The problem of concept drift: definitions and related work URL: https://www.researchgate.net/publication/228723141_The_Problem_of_Concept_Drift_Definitions_and_Related_Work
51. Žliobaitė I, Pechenizkiy M, Gama J. An overview of concept drift applications. In: Big Data Analysis: New Algorithms for a New Society. Switzerland: Springer International Publishing; 2016:91-114.
52. Martin G, Martin P, Hankin C, Darzi A, Kinross J. Cybersecurity and healthcare: how safe are we? Br Med J 2017;358:j3179. [doi: 10.1136/bmj.j3179] [Medline: 28684400]
53. Clarke R, Youngstein T. Cyberattack on Britain's national health service - a wake-up call for modern medicine. N Engl J Med 2017 Aug 3;377(5):409-411. [doi: 10.1056/NEJMp1706754] [Medline: 28591519]
54. Sittig DF, Singh H. A socio-technical approach to preventing, mitigating, and recovering from ransomware attacks. Appl Clin Inform 2016;7(2):624-632 [FREE Full text] [doi: 10.4338/ACI-2016-04-SOA-0064] [Medline: 27437066]
55. Youssef AE. A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments. Int J Ambient Syst Appl 2014;2(2):1-11. [doi: 10.5121/ijasa.2014.2201]
56. Jogova M, Shaw J, Jamieson T. The regulatory challenge of mobile health: lessons for Canada. Healthc Policy 2019;14(3):19-28. [doi: 10.12927/hcpol.2019.25795] [Medline: 31017863]
57. Torbica A, Cappellaro G. Uptake and diffusion of medical technology innovation in Europe: what role for funding and procurement policies? J Med Market 2010 Jan;10(1):61-69. [doi: 10.1057/jmm.2009.48]
58. Allen B, Wade E, Dickinson H. Bridging the divide - commercial procurement and supply chain management: are there lessons for health care commissioning in England? J Public Procure 2009 Mar;9(1):79-108. [doi: 10.1108/JOPP-09-01-2009-B003]
59. Perrow C. Normal Accidents: Living With High Risk Technologies. Princeton, NJ: Princeton University Press; 2011.

60. Vogus TJ, Welbourne TM. Structuring for high reliability: HR practices and mindful processes in reliability-seeking organizations. J Organ Behav 2003 Nov;24(7):877-903. [doi: 10.1002/job.221]
61. Zuboff S. Big other: surveillance capitalism and the prospects of an information civilization. J Inf Technol 2015;30(1):75-89. [doi: 10.1057/jit.2015.5]
62. CB Insights. 2018. AI In Healthcare Heatmap: From Diagnostics To Drug Discovery, Deals Heats Up URL: https://www.cbinsights.com/research/artificial-intelligence-healthcare-heatmap-expert-intelligence/
63. Scott WR. Institutions and Organizations: Ideas, Interests, and Identities. New York: Sage Publications; 2013.
64. North DC. Institutions. J Econ Perspect 1991;5(1):97-112. [doi: 10.1257/jep.5.1.97]
65. Battilana J, Leca B, Boxenbaum E. How actors change institutions: towards a theory of institutional entrepreneurship. Acad Manag Ann 2009 Jan;3(1):65-107. [doi: 10.1080/19416520903053598]

## Abbreviations

**AI:** artificial intelligence
**GPT:** general purpose technology
**ML:** machine learning
**NASSS:** Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability

*Edited by G Eysenbach; submitted 07.02.19; peer-reviewed by KL Ong, S Chen, S Zheng; comments to author 27.04.19; revised version received 16.05.19; accepted 31.05.19; published 10.07.19.*

*Please cite as:*
*Shaw J, Rudzicz F, Jamieson T, Goldfarb A*
*Artificial Intelligence and the Implementation Challenge*
*J Med Internet Res 2019;21(7):e13659*
*URL: https://www.jmir.org/2019/7/e13659/*
*doi:10.2196/13659*
*PMID:31293245*

Original Paper

# Development and Evaluation of ClientBot: Patient-Like Conversational Agent to Train Basic Counseling Skills

Michael J Tanana[1], PhD; Christina S Soma[2], MEd; Vivek Srikumar[3], PhD; David C Atkins[4], PhD; Zac E Imel[2], PhD

[1]Social Research Institute, College of Social Work, University of Utah, Salt Lake City, UT, United States

[2]College of Education, University of Utah, Salt Lake City, UT, United States

[3]School of Computing, University of Utah, Salt Lake City, UT, United States

[4]Psychiatry and Behavioral Science, University of Washington, Seattle, UT, United States

**Corresponding Author:**
Michael J Tanana, PhD
Social Research Institute
College of Social Work
University of Utah
395 South 1500 East #111
Salt Lake City, UT,
United States
Phone: 1 801 585 9228
Email: michael.tanana@utah.edu

## Abstract

**Background:** Training therapists is both expensive and time-consuming. Degree–based training can require tens of thousands of dollars and hundreds of hours of expert instruction. Counseling skills practice often involves role-plays, standardized patients, or practice with real clients. Performance–based feedback is critical for skill development and expertise, but trainee therapists often receive minimal and subjective feedback, which is distal to their skill practice.

**Objective:** In this study, we developed and evaluated a patient-like neural conversational agent, which provides real-time feedback to trainees via chat–based interaction.

**Methods:** The text–based conversational agent was trained on an archive of 2354 psychotherapy transcripts and provided specific feedback on the use of basic interviewing and counseling skills (ie, open questions and reflections—summary statements of what a client has said). A total of 151 nontherapists were randomized to either (1) immediate feedback on their use of open questions and reflections during practice session with ClientBot or (2) initial education and encouragement on the skills.

**Results:** Participants in the ClientBot condition used 91% (21.4/11.2) more reflections during practice with feedback ($P$<.001) and 76% (14.1/8) more reflections after feedback was removed ($P$<.001) relative to the control group. The treatment group used more open questions during training but not after feedback was removed, suggesting that certain skills may not improve with performance–based feedback. Finally, after feedback was removed, the ClientBot group used 31% (32.5/24.7) more listening skills overall ($P$<.001).

**Conclusions:** This proof-of-concept study demonstrates that practice and feedback can improve trainee use of basic counseling skills.

***(J Med Internet Res 2019;21(7):e12529)*** doi:10.2196/12529

**KEYWORDS**

psychotherapy training; interactive learning; conversational agents; deep learning

## Introduction

### Mental Health Treatment in the United States

In 2014, 43 million adults (18% of the population) in the United States were diagnosed with a mental illness and 21 million Americans with a substance use disorder [1]. Despite a severe need for treatment, less than half of those individuals received mental health services [2]. There is a severe shortage of mental health providers [3], and thus, seeking care can involve many calls to providers [4] and wait times that are longer than the duration of many acute mental health episodes [5]. Moreover, training licensed master's- or doctoral-level psychotherapists is an expensive and time-consuming process. There is incredible societal need to reduce the burden of mental illness and addiction, but a limited workforce and barriers to the rapid and effective training of providers create challenges for addressing these concerns.

## Psychotherapy Training

Psychotherapy training typically includes didactic classroom-based academic instruction, skills practice via role-plays with peers, viewing and discussing recordings of experienced psychotherapists, and clinical supervision, with supervision playing the most significant role [6,7]. Ideally, supervision includes review of recorded sessions and specific performance-based feedback from a competent supervisor. Gold-standard training for licensed therapists includes a workshop-based introduction to a treatment approach and then posttraining support, including coaching and performance-based feedback via a behavioral coding fidelity measure. There is strong evidence that providing ongoing performance-based feedback via behavioral coding to therapists results in skills acquisition and retention (eg, [8]). However, this process is slow and labor intensive (eg, in some cases 4 or 5 times the length of the session) [9]. Consequently, specific and objective feedback based on behavioral coding is rarely used in training.

Even when feedback is available, it usually occurs long after the actual performance of the therapy and is generally vague [10]. Supervision and training primarily rely on the therapist's self-report of what occurred in client sessions [11]. Supervision can be general and highly selective in nature, as opposed to targeting specific behaviors [6,7]. The Beutler study [6] observed, "trainees are provided with suggestions for addressing crises and major problems too late to benefit the patient, and even then, the supervision is typically poorly focused and provides few means to assess improvement." For example, training in basic interviewing/active listening sills (eg, open questions and reflections) is foundational to training in mental health counseling and much of the medical field generally [12,13], and Motivational Interviewing (MI), which is partly based on the use of these skills, is a widely used evidence-based treatment [14]. However, treatments such as MI typically rely on workshops where opportunities for practice and feedback are fairly limited.

Research from cognitive science suggests delayed, nonspecific feedback is not sufficient to promote learning and develop expertise [15]. It has long been established that immediate feedback on specific behaviors is an optimal part of a training regimen with large, positive effects on learning [16]. When this feedback is done correctly, it can outweigh other powerful effects on learning, such as cognitive ability and socioeconomic influences [15]. Typical psychotherapy training and supervision does not meet these optimal conditions, and trainee therapists rarely receive feedback as they are performing the skills themselves.

Another practical difficulty with training therapists is to provide initial skills practice without relying on actual clients. Standardized patients, who are actors that simulate clients and their problems [17], reduce the risk of harming clients with untrained therapists, but they can be expensive or difficult to train. Screening for low-severity clients is another alternative, though they can be difficult and time-consuming to recruit (requiring senior staff time to screen and supervise). Despite best efforts and screening, these clients may ultimately reveal severe mental health concerns. In summary, from the view of the cognitive science literature, ideal psychotherapy training would include many opportunities to practice, with immediate performance-based feedback. However, many practical barriers currently prevent psychotherapy training from meeting these conditions.

## Machine Learning and Psychotherapy

The field of computer science, and specifically machine learning, may provide potential solutions to availability of clients and lack of immediate feedback. Machine learning describes the process of creating algorithms through which a computer continues to learn from the algorithm without continued human interaction [18]. Recent developments in the field of machine learning and artificial intelligence may present solutions to standardizing and scaling up psychotherapy training [19]. Natural Language Processing (NLP) is a subgroup of machine learning, whereby the goal is to "learn, understand, and produce human language content" computationally [20], and recent work has begun to apply NLP to the training of mental health providers.

### *Natural Language Processing–Based Feedback*

First, improvements in NLP have allowed computational models to replicate behavioral coding evaluations of psychotherapy that typically require trained human evaluators [21,22]. Currently, NLP models are able to identify key aspects of MI [23] (eg, questions and reflections)—an evidence-based psychotherapy for substance abuse problems—with similar accuracy to human raters [24]. This new technology allows for the possibility of a computer giving immediate feedback that would not be possible with human raters [25,26]. These new technologies create an opportunity to provide trainees with more rapid feedback that does not rely on resource-intensive human supervision.

### *Neural Conversational Agents as Standardized Patients*

In addition to NLP-based evaluation of therapy, conversational agents may provide a computerized environment for practicing skills, potentially replacing standardized patients in some context. Conversational agents are computer programs that are intended to interact with a real person using language [27]. Early conversational agents relied on rule-based programming with long lists of if-then rules, which limits the ability to adapt a conversational model to a new domain. A recent, major innovation in computer-modeled conversational agents were algorithms that could generate plausible speech without relying on human-generated rules (ie, neural conversational models) that self-teach how to engage in dialogue, learning from a large corpus of examples (eg, recursive neural networks) [28]. Conversational agents have been utilized for training in the medical field [29] but have not yet been applied to training in psychotherapy.

Although conversational agents have not been used in psychotherapy training, there have been attempts to utilize technology to support skills practice and assessment. For example, the Rosengren study [30] created a system, whereby therapists were presented with standardized patient video vignettes and were asked to respond using MI skills (in written form). Their responses were later scored by human raters for MI fidelity [31]. This method has the advantage of providing a truly standardized patient; however, the patient did not respond to the therapist, preventing a more natural clinical exchange. In addition, the system requires a human to score the responses, delaying the receipt of feedback. The Baer study [32] developed a similar system, whereby therapists were presented with video clips and asked to respond as their therapist. Again, responses were scored later by a human for adherence to MI best practices. Thus, neither system has the ability to provide feedback immediately after each therapist response. New NLP models have created the opportunity for simulating a standardized patient without the cost of recruiting and training human patients.

### This Study

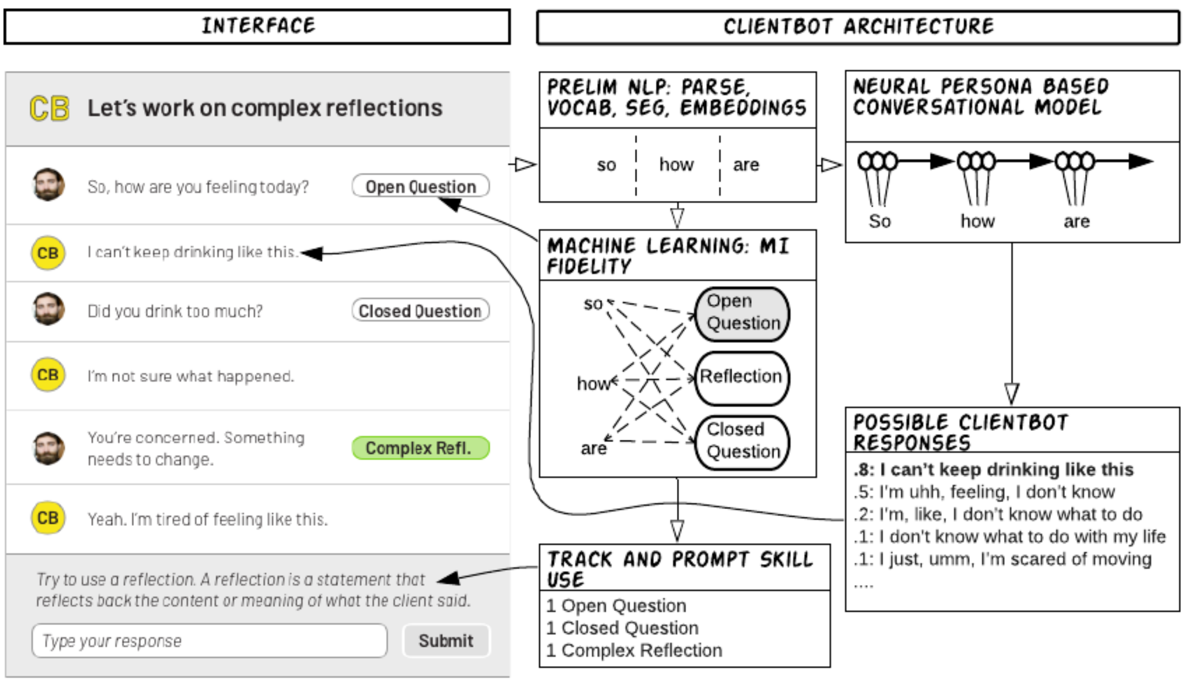
To address the challenges related to the need for scale and immediacy in training new skills in psychotherapy, we developed and evaluated a Web–based system that uses machine learning–based feedback for training 2 specific counseling skills: open questions and reflections. The feedback is embedded into a text–based neural conversational agent, developed to be a standardized patient. Thus, the skills training relied on an automated standardized patient—ClientBot—which provided real-time feedback to trainees on their utilization of specific counseling skills. We randomized nontherapist participants to receive real-time feedback on skill use (or not) and hypothesized that participants in the feedback condition would use more desirable counseling skills (ie, open questions and reflections) after training has ended than in the no feedback condition.

## Methods

### ClientBot Development and Overview

To a trainee participant, the ClientBot platform appears like a standard chat interface, much like what a person might use if they were chatting on the Web with a friend or having a short message service (SMS) text message conversation on a mobile phone (see Figure 1). The key difference is that in this chat platform, the beginning therapist is interacting with a simulated patient, which responds to the trainee's statements using neural network conversational models (described below). Although the trainee interacts with the simulated patient, ClientBot provides feedback on the individual's chat responses, tailored to the skill they are currently practicing—either open questions or reflections. In the following sections, we describe the underlying models and development of the neural conversational model and machine learning–based feedback.

**Figure 1.** Client bot architecture.




#### *Long Short-Term Memory Cell*

The simulated patient in ClientBot is a combination of 2 neural network systems with different strengths and limitations. Both rely on a long short-term memory (LSTM) cell, which is a variation on a traditional neural network sigmoidal unit that solves challenges in learning sequences with long-term dependencies. Similar types of models have been used for quite some time [33], but unfortunately, suffered from the "vanishing gradient problem" [34]. LSTM models solved many of these problems related to learning long sequences [35]. Here, we introduce the 2 LSTM recurrent neural networks underlying ClientBot and provide sample interactions.

The first model is a sequence to sequence model, similar to the model from the Vinyals study [28]. This class of models use an LSTM encoder network to read the input statement, and then have a separate, linked LSTM that decodes the statement into a response. Vinyals et al [28] adapted these models that were initially used for machine translation to be used in dialogue generation. The intuition is that dialogue can be thought of as a similar NLP problem to decoding a French sentence to an English sentence. This model is trained on a collection of English movie transcriptions (ie, open-subs dataset) [36] and 2354 psychotherapy transcripts published by Alexander Street Press, which is available for download with a library subscription [37]. The model was trained using adaptive gradient descent with a learning rate of .01 with gradient clipping of 5 [38], using a vocabulary of 15 thousand words and 1 layer of 3000 LSTM cells. This model achieved a perplexity of 9.06 on a random 5% of data that was not used during training. (Perplexity is a measure of how well the model predicted the next word in a sentence given the previous words, with lower values indicating better fit). When interacting with the model, it uses beam search decoding (with a beam size of 10). Each therapist statement was entered into the encoder, and the client statement was used as the output or criterion for the decoder model. The broad goal of choosing these 2 training sets was to provide the model conversational text, and in particular, the Alexander Street Press transcripts provided specificity of the content and nature of therapeutic conversation. Textbox 1 (Sequence to Sequence or seq2seq) shows example interactions from the model. These examples demonstrate that this model provides brief but plausible responses to questions, which are often indicative of “small talk.” However, this is also its limitation: It does poorly at providing longer responses, which would be typical from a client in a psychotherapy session. Another limitation is that it responds with the phrase “I don’t know” relatively frequently. Finally, because it is partly trained with a corpus of movie transcripts, it responds in ways that would be contextually rare in psychotherapy, such as “I love you.” As a result, after training the model, we undersampled responses with “I don’t know” and did not allow responses that contain “I love you.” These were the only manually developed adaptations for these models. Models were selected based on their accuracy predicting responses in a random 5% of the dataset using perplexity, a standard measurement of how well a language model fits the data. The seq2seq model achieved a perplexity of 9.06 (lower is better) on the test set of examples.

The second model underlying ClientBot is an LSTM model that is only trained on the Alexander Street Press psychotherapy dataset. This model inserted tags to indicate when a speaker was changing from a patient to a therapist and vice versa: (“<p>” and “<t>”). This model is a simpler variation of the seq2seq model. Instead of encoding and decoding, it is just predicting long sequences of words (50 word sequences for our training). The tags indicate the speaker changes and roles. This model used a learning rate of .05 and 1024 LSTM model hidden units in a single layer. On a 5% test set of the Alexander Street Press data, this model achieved a perplexity of 38.01 (note: perplexity is not comparable across different datasets and different vocabularies but rather used to compare models within a given dataset). The decoding for this model randomly sampled at each time step, instead of taking the best next word. The advantage of this strategy is that it increases the likelihood of seeing longer responses from the model, and it also makes the responses nondeterministic. Similar to the previous model, “I don’t know” is undersampled because of its frequency.

Textbox 2 (LSTM model) shows an example interaction with this second model simulating the role of the client. Compared with the first model, this model produces longer talk turns and is more likely to interject psychotherapy like topical matter.

For the ClientBot used in the training, we combined the 2 models by using the first model (sequence to sequence) for the first 5 responses with a trainee, after which the system switches to the second model (simple LSTM model). The reason for starting with the sequence to sequence model is that it generates conversation that often begins a session (eg, generic chitchat and general questions and responses). However, as the session progresses, these short responses are less useful than those generated in the second model, which is prone to discuss problems and symptoms at length.

**Textbox 1.** Example interaction with the Seq2Seq conversational model.

1. Human: Hi

   Top 5 responses:

   - Hi.
   - How are you?
   - What are you doing here?
   - What are you doing?
   - How are you doing?

2. Human: Where are you from?

   Top 5 responses:

   - I don't know
   - I'm from Ohio
   - From the north
   - I don't know…
   - I don't know…but I don't know

3. Human: How old are you?

   Top 5 responses:

   - <number>
   - I'm <number>
   - I don't know
   - <number>?
   - <number>, <number>

4. Human: Goodbye

   Top 5 responses:

   - Goodbye
   - Goodbye.
   - What are you doing?
   - Goodbye...
   - What are you doing here?

**Textbox 2.** Example interaction with the Simple long short-term memory conversational model.

Model: I got a new prescription

Human: what prescription are you on

Model: well, I guess it was just like, I have some.

Human: you have some meds?

Model: yes.

Human: what else do you take?

Model: I take, uh, I take <number> pill, and then I take <number>. You know, and I think it was just one of those things.

Human: How do they make you feel?

Model: oh, my gosh.

Human: Good?

Model: yeah.

### *Machine Learning Prediction of Basic Counseling Skills*

Another central feature of ClientBot is the ability to give immediate, performance–based feedback. In the present case, that feedback takes the shape of classifying each response from the trainee into basic counseling skills. Labeling of these skills was facilitated through a measure of MI, which includes specific behavior counts of several active listening skills. In this study, we focused on open versus closed questions and reflections. To do this, we utilized a discrete sentence feature machine-learning model that provided labels for each trainee statement [24]. The model was trained on a large set of transcripts from a psychotherapy dissemination study that focused on training counselors to do MI. Each transcript was rated using the MI skills code (MISC) [39] (see below in Measures section). The kappa scores for the machine-human agreement ranged between .39 and .79 and were close to human reliability on the codes of interest (open question, closed question, reflections) [24].

ClientBot was run on a server using Torch7 [40] with an Nvidia 980ti graphics processing unit (GPU), which drastically increased the speed of inference (an average example using 1 beam took 30 ms to run on the GPU. On a 4 core CPU, the same example took 500 ms). The time the system takes to identify each statement varies depending on the length of the utterance but was generally less than 1 second. Responses from the conversational agent took between 300 ms to 1 second, depending on which model was used. Participants generally commented that the system responded in a timely manner.

## Participants

For this proof-of-concept study, 151 nontherapists were recruited as participants to assess the effect of the interface on a population with no formal training in counseling. This population ensured that the participants are very unlikely to have been exposed to formal training in counseling skills previously. Participants for this study were recruited from Amazon Mechanical Turk (MTurk) [41]. We limited our sample to either "master workers," who are workers that have a track record of high accuracy on the tasks on which they have worked in the past or workers with at least 10,000 approved jobs and a 95% overall approval rate. We also limited the sample to US residents who spoke US English and excluded participants under the age of 18 years. The amount that workers were paid depended on the demand for work at the time that they enrolled, which varied between US $3 and US $3.50 for each participant. Each potential participant was recruited to "practice your listening skills," inviting interested people "...to chat with a simulated person for 20 min and practice their listening skills." Participants then completed a short (10 question) survey when they were done.

## Procedure

Interested participants were directed to a page where they read the consent form. If they agreed to participate they were then randomized into 1 of 2 conditions. Both conditions were given a brief introduction to "listening skills," focused on reflections and open questions including various examples. At the end, participants took a 3-question quiz to ensure that participants understood what skills they were supposed to be practicing. Users were allowed to go back and read this introduction after reading the questions.

Both conditions included discrete phases focused on different skills, as shown in Figure 2. In the control condition, after reading the introduction, users began to interact with the simulated client. These users were prompted with skill-specific introductory prompts (eg, "now practice open questions," and "now practice reflections") but received no feedback on their interactions with the simulated client. Participants in the treatment condition read the same introductory statement, prompts, and training as the control condition ("now practice open questions"). If a user was not responding with, for example, enough reflections during the reflection training section, the system prompted them to practice more reflections and give examples similar to the introduction. All participants then had a 5-min test phase where all prompts and feedback are removed from the system ("For the last 5 minutes, show us your best listening skills"). After interacting with ClientBot, participants completed several questionnaires (see Measures).

Figure 2 shows the progression through the stages of the curriculum. Boxes in the middle of the figure represent components that both the treatment and control groups received. Items on the right side of the figure show components that only the treatment group received.

## Measures

### *Open Questions and Reflections*

As noted above, ClientBot includes a machine-learning engine trained to identify categories of basic counseling skills assessed by a standard measure of MI fidelity [24]—the MISC [39]. We used models from the methods described in the study by Tanana et al [24]. These MISC identification models could identify open and closed questions and reflections on a test set with similar performance to human-human reliability (see [24] for full table of results). To track the success of training, the number of open question and reflection codes are tabulated and divided by the total number of utterances, yielding percentages of each type of statement. The primary outcomes during the training session itself were percentages of reflections and open questions.

### *Posttest Fixed Responses*

The primary outcome to measure changes in MI desirable behaviors was the use of open questions and reflections during the interactive session with the simulated client. However, to guard against the possibility that the simulated person could create a self-reinforcing loop, exaggerating group differences, the users were also asked to respond to 5 standardized client responses during the posttest, using the skills they had learned (these prompts were exactly the same for all participants). After the completion of the curriculum, participants were asked to respond to 5 example client statements on a survey using the skills that they had learned. Their responses were coded using the automate fidelity system described above.

### *Satisfaction*

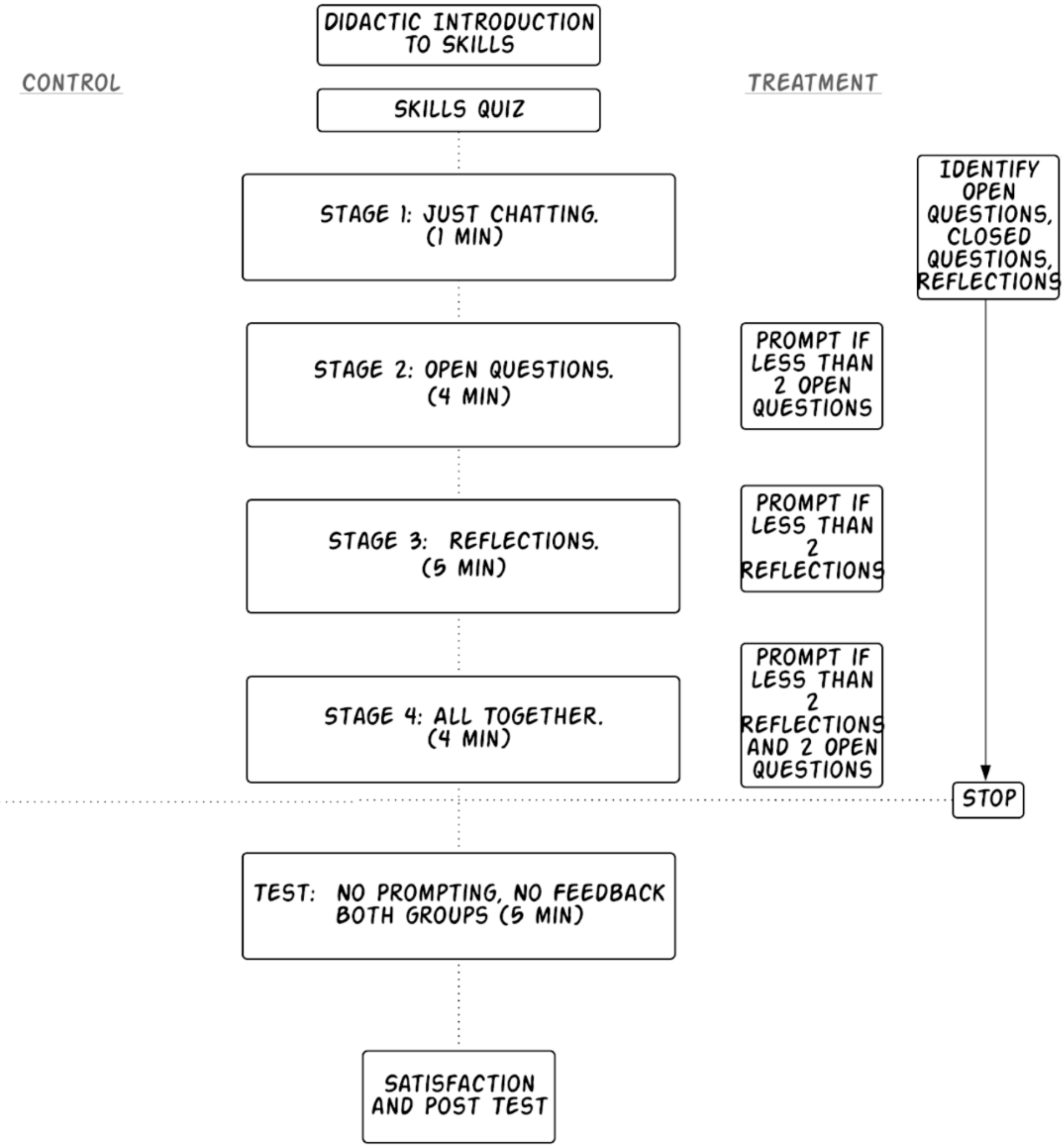
We measured 2 types of satisfaction: (1) Satisfaction with the ClientBot system in general and (2) Satisfaction with the ClientBot simulated client. The first included questions such as

"I thought the system gave me useful information" and "I would use the system again." The second construct asked questions such as "The simulated person was interesting to talk to" and "I found the simulated person to be tedious to interact with." All questions had the responses strongly agree, agree, disagree, and strongly disagree. Satisfaction scores were coded from 1 to 5 with 1 indicating strong dissatisfaction, 3 indicating a neutral response, and 5 indicating strong satisfaction.

## Data Analyses

Hypotheses 1 and 2 were tested by comparing the percentage of statements that were open questions and reflections between the treatment and control group (1) during the training phase and (2) during the test stage of the training (see Figure 2). The comparisons were made using a Wilcoxon rank test because the outcome was a transformation of a count variable.

**Figure 2.** Client bot curriculum and design.



## Results

### Performance

After consenting to the study, 22 users (12.1%) did not complete all phases of the interaction with the simulated client. This typically happened after only a few talk turns (median=9). There was no statistical difference in the rate of dropout between the treatment (15%) and the control group (10%; $X^2_1$ =0.51; *P*=.47). These users were excluded from all analyses.

A total of 151 participants completed the study (73 participants in the treatment group and 78 in the control group). The characteristics of the participants can be seen in Table 1. The sample was relatively balanced for males and females (53.6% male) and contained 25% non-white participants. The educational backgrounds were diverse as well, with 42% having a bachelor's degree or higher.

**Table 1.** Demographics.

| Identity status | Control, n (%) | Treatment, n (%) |
|---|---|---|
| **Race/ethnicity** | | |
| Native American | 1 (1.3) | 0 (0) |
| Asian | 3 (3.9) | 6 (8.3) |
| Black/African American | 7 (9.1) | 8 (11.1) |
| Hispanic/Latino | 6 (7.8) | 3 (4.2) |
| Multiracial | 3 (3.9) | 0 (0) |
| White or not Hispanic | 57 (74) | 54 (75) |
| **Education** | | |
| High school or less | 11 (14.3) | 14 (19.4) |
| Some college | 32 (41.6) | 25 (34.8) |
| Bachelor's degree or more | 34 (44.2) | 33 (45.8) |
| **Gender** | | |
| Female | 38 (49.4) | 31 (43.1) |
| Male | 39 (50.6) | 41 (56.9) |

At the outset of the study, after reading the initial introduction to open questions and reflections, participants answered 3 questions to test their understanding (Note that participants were able to go back and reread the introduction while answering these questions). The purpose of these questions was primarily a minimal validation check that participants were engaged in the task. A total of 96.6% of participants correctly answered the question about open questions, 95.3% the question about reflections, and 95.3% the question about the purpose of the study.

To verify that participants were not using a similar response repeatedly during their interactions with ClientBot, the percentage of unique utterances was estimated, with average of 98.5% unique utterances in treatment group and average of 97.3% unique utterances in control. This result indicates that very few participants could have repeated successful responses as a way of artificially inflating their performance or nominally completing the task without engaging in a meaningful way.

Performance of listening skills was assessed at 6 different time points during the study: before training began, during open question training, during reflection training, during combined reflection and open question training, after feedback was removed, and performance on fixed prompts on the posttest (see Tables 2 and 3). There were no differences in skill performance before training began on either open questions or reflections. During open question training, the treatment group used significantly more open questions than the control group (30.4% compared with 22.4%). During the reflection training, the treatment group used significantly more reflections than the control group (21.4% compared with 11.2%). During the combined training section, the treatment group used significantly more reflections than the control group (15.8% compared with 9.3%), but both groups used similar rates of open questions. After feedback was removed, the treatment group continued to use more reflections than the control group (14.1% compared with 8%), and both groups used similar rates of open questions.

Results in the posttraining assessment followed the same pattern as the responses with the simulated client. There were no significant differences in the use of open questions between the treatment and control group ($W_{149}$=3040; *P*=.29). However, the treatment group used significantly more reflections than the control group ($W_{149}$=1800; *P*<.01; *d*=0.58).

**Table 2.** Sample sizes for results.

| Participant sample | Statistics, n | |
|---|---|---|
| | Control | Treatment |
| Before training | 86 | 86 |
| Open Question training | 82 | 82 |
| Reflection training | 80 | 75 |
| Training both | 79 | 74 |
| Test (feedback removed) | 79 | 74 |
| Fixed responses (posttest) | 78 | 71 |

**Table 3.** Results of the assessment.

| Task and Time Frame, (pre, training, post) | Open questions | | | Reflections | | | Reflection or open questions | | |
|---|---|---|---|---|---|---|---|---|---|
| | Control | Treatment | *P* value | Control | Treatment | *P* value | Control | Treatment | *P* value |
| Before training | 25.5 | 23.8 | .97 | 8.6 | 8.6 | .81 | 34.1 | 32.4 | .77 |
| Open Question training | 22.4 | 30.4 | <.001 | 6.3 | 5 | .14 | 28.6 | 35.2 | <.001 |
| Reflection training | 15.6 | 11 | .003 | 11.2 | 21.4 | <.001 | 26.8 | 32.4 | .002 |
| Training both | 18.4 | 20.9 | .07 | 9.3 | 15.8 | <.001 | 27.7 | 36.8 | <.001 |
| Test (feedback removed) | 16.7 | 18.3 | .16 | 8 | 14.1 | <.001 | 24.7 | 32.5 | <.001 |
| Fixed responses (posttest) | 40 | 35.1 | .29 | 18.2 | 34.6 | <.001 | 58.2 | 60.7 | .01 |

### Satisfaction

There were no significant differences between groups on overall satisfaction, satisfaction with the conversational agent, or satisfaction with the system in specific. Satisfaction was not significantly different from neutral (2.85; $t_{148}$=–1.91; *P*=.06), with most of this effect attributed to dissatisfaction with the simulated person (2.39; $t_{148}$=–6.28; *P*<.01) and a nearly neutral response to the system overall (3.02; $t_{148}$=0.28; *P*=.77). There was no difference in system satisfaction by group ($t_{141}$=0.021; *P*=.98), simulated person satisfaction by group ($t_{146}$=0.552; *P*=.58), or overall satisfaction by group ($t_{144}$=0.201; *P*=.84).

The majority of respondents said that system was not boring (70%) and that they thought the system gave useful information (75%). The participants were split on whether or not they would use the system again, with 46% reporting that they would. Only 35% of users thought that the simulated person was interesting to talk with, and a majority also thought that interacting with the simulated person was tedious (73%).

## Discussion

### Results Summary and Inferences

This study investigated a new methodology for teaching active listening skills to an untrained population using a computerized simulated patient, and automated feedback, that could all be delivered without experts supervising each individual directly. These initial results indicate that an untrained population can improve specific types of listening skills very quickly (in 20 min). The treatment group in this study had higher rates of reflections, and maintained their increased rate of reflections, even after the feedback and prompts went away. The control group showed an initial propensity to use open questions, even without feedback, but demonstrated a steady decay of open questions over time.

Surprisingly, there were no treatment effects for open questions. That is, only a brief introduction on open questions and some practice elicited use of open questions. These types of utterances can be produced by simply using a set of sentence stems ("How," "Why," "What"). As a result, an open question may be an easier skill to learn than a reflection, and less feedback is required. In contrast to open questions, reflections notably increased with feedback. A reflection involves listening to another person and responding with a summary or refrain of what that person has been trying to express. For example, if a client discussed concerns related to waking up with headaches and often missing work because of parties, a reflection might be to say, "so it sounds like you are worried that drinking is getting in the way of the things you would like to be doing in life." For individuals who have never been exposed to MI or basic counseling skills training, reflections may be a less intuitive skill than an open question. In addition, during the survey following the training, some of the participants noted that they found it much more difficult to produce reflections than open questions.

This study primarily focused on the acquisition of 2 basic listening skills; however, there is some tentative evidence for the durability of the gains. After training, both groups took a satisfaction and demographic survey, and then were surprised with 5 more client statements that they were asked to respond to using the listening skills they had learned during the training. Although not a formal distraction task, the treatment group retained the skills from the training relative to the control; a promising result for later research into the durability of these gains.

Satisfaction is a secondary outcome compared with changes in the practice of skills but a potentially important factor for dissemination of a system such as the one tested in this study. Users in this study had a negative view of the simulated person and a neutral view of the system as a whole. It should be noted that there was no comparison with a more traditional curriculum that consisted purely of written material, and as a result, the view of satisfaction should be interpreted with caution. Users may have enjoyed this study's experience more than the latter. However, the results suggest that efforts should be made to improve the user experience.

### Limitations and Future Directions

One important limitation of this study is that its participants were workers from Amazon MTurk. This is clearly a different population from students who might be starting a mental health training program (eg, social work, psychology, and psychiatry). However, this limitation presents some advantages; notably, these results should generalize to a wider population than just individuals who could be accepted to a counseling graduate program. One of the major limitations of typical psychology research is that it often relies heavily on undergraduate college populations, often limiting the generalization of this research. The MTurk population tends to more closely represent the US population and is much more diverse than the typical sample of undergraduate students [42]. However, further research should be conducted to verify that this type of approach can also benefit the population that does enter a graduate program in psychology.

This study primarily tested differences in acquisition of open questions and reflections and did not test differences in retention or transfer of learning. The Schmidt study [43] has pointed out that treatment differences in acquisition do not necessarily have an impact on retention and transfer. Future research should follow and test participants a week or more after treatment, possibly with multiple administrations of the treatment. Moreover, there should be an investigation into the effects of written training on spoken interactions. This study does not answer the question of how well practicing in a chat forum may transfer to an actual therapy setting. It is possible and even likely that many of the manipulations that have drastic improvements on acquisition may have much lower impacts on retention and transfer of learning.

There was a general sense among the participants of the study that the simulated patient was not a realistic substitute for another human. The computerized dialogue model could sometimes say distracting or irrelevant responses. It is important to note that these models were trained on a relatively small sample of dialogue compared with similar models published in the literature. For example, the Vinyals study [28] used 62 million training examples, whereas the corpus of psychotherapy transcripts used in this study only has 514,118 examples. Moreover, the dialogue in which these models trained was transcribed from actual spoken interactions, which tend to be filled with disfluencies and often trail off. Many of the original transcripts can be hard to understand for these reasons, so it is not surprising that the model trained on these transcripts can occasionally respond in a way that seems out of place. Despite the user's dissatisfaction with the simulated person, the conversational agent did create thousands of novel utterances that the participant could use to practice their listening skills.

The bot performance might be improved by utilizing chat transcripts from Web–based therapy or crisis interventions via SMS text messages (eg, Crisis Text Line, Talk Space, and 7-Cups). Other sources of written text that might be relevant include the Reddit mental health–related forums; however, these function differently than traditional dialogue. This current conversational model is not able to track long-term topical dependencies in a dialogue but rather just attempts to create a likely response to the last talk turn. A more engaging and believable model will benefit from methods that can capture these long-term dependencies in a conversation. However, it may be possible that it is not entirely necessary for ClientBot to fully replicate the experience of talking to another human to provide a useful and satisfactory training experience. Ironically, there is evidence that as the bot begins to further approach a fully human-like presentation, it may become less satisfactory or odd (eg, the uncanny valley).

At a more conceptual level, ClientBot is a technology that is focused on supporting a human's ability to communicate more effectively with other humans. Thus, its use raises fundamental questions about the relationship between humans and machines, or more specifically how humans function in these computer-supported learning environments. For example, it maybe that humans are more apt to trust feedback they receive from a computer (rather than a human) as they see it as more objective [44], even though machine learning–based ratings from a computer are prone to bias and error in a way commensurate with the data from which they are trained. [45]. Accordingly, it may be important to adapt future systems to help humans appropriate challenge the evaluations they receive [46]. Ultimately, the improvement of systems such as ClientBot will rely on ongoing "human in the loop" feedback [47], whereby users learn from the system and also provide feedback and insights that serve to make the platform more effective.

One interesting direction for this research is to possibly develop algorithms that can *produce* a sample reflection for a client statement. For example, in this study, when a trainee responds with an open question when they are supposed to be practicing reflections, the system may prompt them to "keep practicing reflections" and give a generic example of a reflection. In contrast, an ideal system may take the last client statement ("I just don't know what to do, my work day never seems to end") and give an example reflection ("you're feeling overwhelmed at your job"). It is not unreasonable to think that this type of model is plausible given the current state of NLP. NLP researchers have become excellent at question answering tasks [48], which is relatively similar to the problem of producing a reflection. Finally, this study examined both feedback on therapist talk turns as well as adaptive prompting. The treatment effects include both of these tools combined. In addition, both the control group and treatment group received prompting and a didactic explanation of the skills they were supposed to be practicing. Each of these components likely has an effect on the outcomes measured in the study. It would be beneficial to break down each of these skills into a separate component. This type of study would require a much larger pool of participants but would contribute useful knowledge about the impact of various training modalities.

## Conclusions

During the course of the last half century, fields such as aviation and medicine have used technology to augment and enhance the capabilities of humans. In contrast, psychotherapy training and practice generally look very similar to the way they did 50 years ago. Moreover, psychotherapy has the additional problem that there is no natural feedback loop providing practitioners with a means to improve over time [49]. This study tested a method for both providing feedback and training that has the possibility to scale beyond the time limits of a single expert trainer. The results show that at least for the population that participated in this study, that this methodology can improve performance of specific listening skills. This type of system presents a promising avenue to improve the scale on which feedback, adherence, and training can be delivered in the field of psychotherapy.

## Conflicts of Interest

Authors MJT, ZEI, and DCA are co-founders and part owners of Lyssn.io, a company that uses speech and signal processing to provide feedback on psychotherapy skills.

## References


1. Substance Abuse and Mental Health Services Administration (US). 2018. 2015 National Survey on Drug Use and Health: Summary of the Effects of the 2015 NSDUH Questionnaire Redesign: Implications for Data Users. Rockville (MD) URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2015/NSDUH-DetTabs-2015/NSDUH-DetTabs-2015.htm [accessed 2019-05-11] [WebCite Cache ID 78ICth8Et]
2. Han B, Hedden SL, Lipari R, Copello EA, Kroutil LA. Substance Abuse and Mental Health Services Administration. 2015. Receipt of services for behavioral health problems: Results from the 2014 National Survey on Drug Use and Health URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DR-FRR3-2014/NSDUH-DR-FRR3-2014/NSDUH-DR-FRR3-2014.htm [accessed 2019-05-11] [WebCite Cache ID 78IDN0B2d]
3. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. Psychiatr Serv 2009 Oct;60(10):1323-1328. [Medline: 19797371]
4. Kugelmass H. "Sorry, I'm Not Accepting New Patients": an audit study of access to mental health care. J Health Soc Behav 2016 Dec;57(2):168-183. [doi: 10.1177/0022146516647098] [Medline: 27251890]
5. Gallucci G, Swartz W, Hackerman F. Impact of the wait for an initial appointment on the rate of kept appointments at a mental health center. Psychiatr Serv 2005 Mar;56(3):344-346. [Medline: 15746510]
6. Beutler LE, Harwood TM. Virtual reality in psychotherapy training. J Clin Psychol 2004 Mar;60(3):317-330. [Medline: 14981794]
7. Rønnestad MH, Ladany N. The impact of psychotherapy training: introduction to the special section. Psychother Res 2006 May 1;16(3):261-267. [doi: 10.1080/10503300600612241]
8. Schwalbe CS, Oh HY, Zweben A. Sustaining motivational interviewing: a meta-analysis of training studies. Addiction 2014 Aug;109(8):1287-1294. [doi: 10.1111/add.12558] [Medline: 24661345]
9. Moyers TB, Martin T, Manuel JK, Hendrickson SM, Miller WR. Assessing competence in the use of motivational interviewing. J Subst Abuse Treat 2005 Jan;28(1):19-26. [doi: 10.1016/j.jsat.2004.11.001] [Medline: 15723728]
10. Milne D, Westerman C. Evidence-based clinical supervision: rationale and illustration. Clin Psychol Psychother 2001 Nov;8(6):444-457. [doi: 10.1002/cpp.297]
11. Falender CA, Cornish JAE, Goodyear R, Hatcher R, Kaslow NJ, Leventhal G, et al. Defining competencies in psychology supervision: a consensus statement. J Clin Psychol 2004 Jul;60(7):771-785. [Medline: 15195339]
12. Lloyd M, Bor R, Noble L. Clinical Communication Skills for Medicine. London, UK: Elsevier Health Sciences; 2018.
13. Hill C, O'Brien K. Helping skills: Facilitating exploration, insight, and action. Washington DC: American Psychological Association; 2004.
14. Miller W, Rollnick S. Motivational Interviewing: Helping People Change, 3rd Edition (Applications of Motivational Interviewing). New York: Guilford; 2012.
15. Hattie J, Timperley H. The power of feedback. Rev Edu Res 2016 Nov 16;77(1):81-112. [doi: 10.3102/003465430298487]
16. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull 1996;119(2):254-284. [doi: 10.1037/0033-2909.119.2.254]
17. Klamen DL, Yudkowsky R. Using standardized patients for formative feedback in an introduction to psychotherapy course. Acad Psychiatry 2002;26(3):168-172. [doi: 10.1176/appi.ap.26.3.168] [Medline: 12824134]
18. Samuel AL. Artificial intelligence: a frontier of automation. Ann Am Acad Pol Soc Sci 2016 Sep 8;340(1):10-20. [doi: 10.1177/000271626234000103]
19. Pace B, Tanana M, Xiao B, Dembe A, Soma C, Steyvers M, et al. What about the words? Natural language processing in psychotherapy. Psychother Bull 2016;51(1):17-18 [FREE Full text]

20. Hirschberg J, Manning CD. Advances in natural language processing. Science 2015 Jul 17;349(6245):261-266. [Medline: 26185244]
21. Atkins DC, Steyvers M, Imel ZE, Smyth P. Scaling up the evaluation of psychotherapy: evaluating motivational interviewing fidelity via statistical text classification. Implement Sci 2014 Apr 24:9-49. [Medline: 24758152]
22. Can D, Marín RA, Georgiou PG, Imel ZE, Atkins DC, Narayanan SS. "It sounds like...": a natural language processing approach to detecting counselor reflections in motivational interviewing. J Couns Psychol 2016 Apr;63(3):343-350. [Medline: 26784286]
23. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change. New York City: Guilford Press; 2012.
24. Tanana M, Hallgren KA, Imel ZE, Atkins DC, Srikumar V. A comparison of natural language processing methods for automated coding of motivational interviewing. J Subst Abuse Treat 2016 Dec;65:43-50 [FREE Full text] [doi: 10.1016/j.jsat.2016.01.006] [Medline: 26944234]
25. Xiao B, Imel ZE, Georgiou PG, Atkins DC, Narayanan SS. "Rate My Therapist": automated detection of empathy in drug and alcohol counseling via speech and language processing. PLoS One 2015;10(12):e0143055 [FREE Full text] [doi: 10.1371/journal.pone.0143055] [Medline: 26630392]
26. Xiao B, Huang C, Imel ZE, Atkins DC, Georgiou P, Narayanan SS. A technology prototype system for rating therapist empathy from audio recordings in addiction counseling. PeerJ Comput Sci 2016 Apr 2:e59. [Medline: 28286867]
27. Jurafsky D, Martin JH. An introduction to natural language processing, computational linguistics, and speech recognition. In: Prentice Hall Series in Artificial Intelligence. Upper Saddle River, New Jersey: Prentice Hall, Pearson Educational International; 2009:1-1024.
28. Vinyals O, Le Q. arXiv. 2015. A Neural Conversational Model URL: https://arxiv.org/abs/1506.05869 [accessed 2019-05-11] [WebCite Cache ID 78IOg4gJF]
29. Hubal RC, Kizakevich PN, Guinn CI, Merino KD, West SL. The virtual standardized patient. Simulated patient-practitioner dialog for patient interview training. Stud Health Technol Inform 2000;70:133-138. [Medline: 10977526]
30. Rosengren DB, Hartzler B, Baer JS, Wells EA, Dunn CW. The video assessment of simulated encounters-revised (VASE-R): reliability and validity of a revised measure of motivational interviewing skills. Drug Alcohol Depend 2008 Sep 1;97(1-2):130-138. [Medline: 18499356]
31. Miller WR, Moyers TB, Ernst D, Amrhein P. The University of New Mexico. 2003. Manual for the motivational interviewing skill code (MISC) URL: https://casaa.unm.edu/download/misc.pdf [accessed 2019-05-15] [WebCite Cache ID 78O3xOuw8]
32. Baer JS, Carpenter KM, Beadnell B, Stoner SA, Ingalsbe MH, Hartzler B, et al. Computer Assessment of Simulated Patient Interviews (CASPI): psychometric properties of a web-based system for the assessment of motivational interviewing skills. J Stud Alcohol Drugs 2012 Jan;73(1):154-164 [FREE Full text] [Medline: 22152673]
33. Haykin S. Neural Networks: A Comprehensive Foundation. Upper Saddle River, New Jersey: Prentice Hall; 1999.
34. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 1994;5(2):157-166. [Medline: 18267787]
35. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997:1735-1780. [Medline: 9377276]
36. Open Subtitles - download movie and TV Series subtitles. Oct 17. URL: http://www.opensubtitles.org/en/search/subs [accessed 2019-05-11] [WebCite Cache ID 78IQ2VdkM]
37. Alexander Street. URL: https://alexanderstreet.com/ [accessed 2019-05-11] [WebCite Cache ID 78IQLikP9]
38. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 2011;12:2121-2159 [FREE Full text]
39. Miller WR, Moyers TB, Ernst D, Amrhein P. The University of New Mexico. 2008. Manual for the motivational interviewing skill code version 2.1 (MISC) URL: https://casaa.unm.edu/download/misc.pdf [accessed 2019-05-15] [WebCite Cache ID 78O4LWBZF]
40. Collobert R, Kavukcuoglu K, Farabet C. Torch7: A matlab-like environment for machine learning. 2011 Presented at: NIPS 2011; December 12-17, 2011; Granada, Spain p. 1-6 URL: https://infoscience.epfl.ch/record/192376/files/Collobert_NIPSWORKSHOP_2011.pdf
41. Amazon Mechanical Turk. URL: https://www.mturk.com/ [accessed 2019-05-11] [WebCite Cache ID 78IQuSX2u]
42. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [Medline: 26162106]
43. Schmidt RA, Bjork RA. New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. Psychol Sci 2017 Apr 25;3(4):207-218. [doi: 10.1111/j.1467-9280.1992.tb00029.x]
44. Hirsch T, Soma C, Merced K, Kuo P, Dembe A, Caperton DD, et al. It's hard to argue with a computer: investigating psychotherapists' attitudes towards automated evaluation. DIS (Des Interact Syst Conf) 2018 Jun;2018:559-571 [FREE Full text] [doi: 10.1145/3196709.3196776] [Medline: 30027158]
45. Courtland R. Bias detectives: the researchers striving to make algorithms fair. Nature 2018 Dec;558(7710):357-360. [doi: 10.1038/d41586-018-05469-3] [Medline: 29925973]
46. Hirsch T, Merced K, Narayanan S, Imel ZE, Atkins DC. Designing contestability: interaction design, machine learning, and mental health. DIS (Des Interact Syst Conf) 2017 Jun;2017:95-99 [FREE Full text] [doi: 10.1145/3064663.3064703] [Medline: 28890949]

47. Narayanan S, Georgiou PG. Behavioral signal processing: deriving human behavioral informatics from speech and language: computational techniques are presented to analyze and model expressed and perceived human behavior-variedly characterized as typical, atypical, distressed, and disordered-from speech and language cues and their applications in health, commerce, education, and beyond. Proc IEEE Inst Electr Electron Eng 2013 Feb 7;101(5):1203-1233 [FREE Full text] [doi: 10.1109/JPROC.2012.2236291] [Medline: 24039277]
48. Weston J, Bordes A, Chopra S, Rush AM, van Merriënboer B, Joulin A, et al. arXiv. 2015. Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks URL: https://arxiv.org/abs/1502.05698v10 [accessed 2019-05-15] [WebCite Cache ID 78O4vFxgL]
49. Tracey TJ, Wampold BE, Lichtenberg JW, Goodyear RK. Expertise in psychotherapy: an elusive goal? Am Psychol 2014 Apr;69(3):218-229. [Medline: 24393136]


## Abbreviations

**LSTM:** long short-term memory
**MI:** Motivational Interviewing
**MISC:** motivational interviewing skills code
**MTurk:** Mechanical Turk
**NLP:** Natural Language Processing
**SMS:** short message service


*Edited by G Eysenbach; submitted 17.10.18; peer-reviewed by S Martino, M Roccetti, G Lim, C Lin, A Louren; comments to author 10.01.19; revised version received 06.03.19; accepted 02.04.19; published 15.07.19.*

*Please cite as:*
*Tanana MJ, Soma CS, Srikumar V, Atkins DC, Imel ZE*
*Development and Evaluation of ClientBot: Patient-Like Conversational Agent to Train Basic Counseling Skills*
*J Med Internet Res 2019;21(7):e12529*
*URL: https://www.jmir.org/2019/7/e12529/*
*doi:10.2196/12529*
*PMID:31309929*

Original Paper

# Design Guidelines for a Technology-Enabled Nutrition Education Program to Support Overweight and Obese Adolescents: Qualitative User-Centered Design Study

Cynthia LeRouge[1], MS, PhD; Polina Durneva[1], BA; Savitha Sangameswaran[2], BTECH; Anne-Marie Gloster[3], RD, MPH, PhD

[1]Health Informatics and Analytics Program, Department of Information Systems & Business Analytics, Florida International University, Miami, FL, United States

[2]Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States

[3]Nutritional Sciences Program, School of Public Health, University of Washington, Seattle, WA, United States

**Corresponding Author:**
Cynthia LeRouge, MS, PhD
Health Informatics and Analytics Program
Department of Information Systems & Business Analytics
Florida International University
Ryder Business Building, 2nd Fl
11200 SW 8th St
Miami, FL, 33199
United States
Phone: 1 305 348 4709
Email: clerouge@fiu.edu

## Abstract

**Background:** Childhood overweight and obesity are major health challenges in the United States. One of the recommendations to combat obesity is to maintain a healthy diet, which is often best supported by eating home-cooked meals to control cooking methods, ingredients, and portions. Diet control through home cooking is challenged because of the decline in culinary skills in the population and a paucity of effective culinary nutrition education (CNE) programs. Providing technology-enabled CNE (CNE-tech) to overweight and obese adolescents can equip them with life skills that can assist them in the future. Such skills can facilitate saving money, eating healthier, and creating social environments. In addition, CNE builds cooking confidence and food literacy that in turn can build adolescent self-efficacy, particularly toward managing their health behaviors.

**Objective:** This study aimed to inform functionalities, design requirements, and the context of use for CNE-tech that could enhance overweight and obese adolescents' healthy food literacy, cooking confidence, and general self-efficacy with regard to self-management to ultimately promote healthy lifestyle management.

**Methods:** The design science study was completed in 2 distinct phases engaging overweight and obese adolescents, parents of overweight and obese adolescents, and the health care providers that treat adolescents with these conditions. Phase 2, our primary source of data, involved user-centered design methods including the following: (1) early stage prototype usability analysis, (2) semistructured interviews with 70 overweight or obese adolescents engaged in a healthy behavior program, and (3) semistructured interviews with 10 health care providers. Data were analyzed using constant comparison analysis to identify functionalities, design requirements, and inform the context of use of CNE-tech.

**Results:** Data revealed specific desired functionalities for the CNE-tech related to building cooking skills, populating a healthy recipe database, suggesting healthy alternatives, supporting the construction of a healthy plate, and the ability to share healthy recipes and cooking accomplishments. Moreover, the adolescents provided design requirements pertaining to the presentation (eg, vivid colors, semirealistic images, and cooking sounds), use of multimedia, and gaming. Data further revealed contextual factors, such as shared experiences with family members and enhanced continued use.

**Conclusions:** We demonstrate the potentiality of creating CNE-tech that could effectively lead to better self-care and induce sustainable behavioral change as it facilitates skill building, self-efficacy, and a pathway that enables overweight and obese adolescents to influence cooking habits in their family home and future dwellings. Our CNE-tech–proposed solution aligns with the goals of overweight and obese adolescents and also reflects existing theories about behavioral change.

*(J Med Internet Res 2019;21(7):e14430)* doi:10.2196/14430

**KEYWORDS**

healthy eating; user-computer interface; consumer health informatics; adolescent; obesity; overweight; nutrition; cooking; user-centered design

## Introduction

### Background

Overweight and obesity are major public health issues affecting 12.7 million children and adolescents [1-3] and the overall health care system, resulting in the annual cost of about US $190 billion in the United States [4]. To date, it has proved to be challenging to isolate and study the effects of single factors (genetic and nongenetic) on this condition, as obesity and overweight seem to be caused by the interplay of a variety of factors [5]. Lack of physical activity, unhealthy eating, or the combination of the two is a primary cause of overweight and obesity in children and adolescents [6]. Genetic predisposition and certain social factors (eg, low socioeconomic factors and physical environment) are also associated with different levels of obesity and overweight [7].

Healthy diet and physical activity are crucial to combat these conditions and maintain healthy weight. Studies have shown that enhanced culinary skills could lead to better diet quality over a number of years as more home-cooked meals would be consumed [8]. Consumers in the United States, unfortunately, have been experiencing a decline in culinary skills [9] and been shifting from home-cooked meals to prepared foods. Previous studies explored such shift with respect to wide availability and low prices of convenience foods [10] and consumers' socioeconomic status, culinary experience, free time, and limited culinary abilities [11-15].

Healthy culinary habits and behaviors have been previously explored, defined, and associated with positive health outcomes [16]. Educational programs and health interventions can, therefore, be effective and important for promoting healthy eating. Traditional nutrition education programs have, however, been shown to be ineffective in the past because of the content taught in such classes [17,18]. These programs focus primarily on educating people about food characteristics and potential medical disorders associated with over or underconsumption of nutrients [19,20] and do not provide typical consumers with skills necessary to prepare and select healthy meal options. Culinary nutrition education (CNE), on the contrary, can be more advantageous by bringing knowledge of healthy eating into action.

CNE programs teach students how to adequately nourish themselves through food, convert raw ingredients into edible healthy dishes, and maintain healthy diet habits in the future [21]. Through CNE, adolescents benefit by increasing food literacy and cooking confidence [22], which have been described as inadequate among the youth [23]. Food literacy can be defined as "the ability to make healthy food choices by having the skills and knowledge necessary to buy, grow, and cook food with implications for improving health" [24]. Cooking confidence refers to confidence in cooking certain meals, implementing various cooking techniques, and following recipes [25]. In addition to potential health benefits, increased cooking confidence and food literacy can build adolescents' self-efficacy [22] to maintain healthy eating behaviors. Knowledge gained through cooking can also help adolescents make informed less calorie-dense choices when choosing to prepare a recipe or when eating at a restaurant [26]. In addition, by influencing meal choices and taking over cooking responsibilities, adolescents may serve as catalysts for healthy eating in their families. Adolescents with high food literacy can build stronger connections with their families, according to Utter et al [27].

Despite their obvious benefits, nutritional health intervention programs are not widely available for US adolescents [28]. One of the ways to address such problem is to integrate lifestyle intervention through information and communication technologies (ICTs) that can provide tailored education in a cost-effective manner. A recent systematic review on the use of nutritional interventions for adolescents using ICTs indicated that game-based ICTs, which focused on promoting healthy habits of adolescents, tended to be quite efficient in their purposes and that "long-term interventions for adolescents that make use of frequent exposure to technological resources, and that have a theoretical component aimed at a single health behavior change, tend to be more successful" [29]. Regarding specific ICTs, improvements in lifestyle habits, such as smoking cessation, have been shown to be positively associated with an increase in mobile text messaging, a type of ICT intervention [30]. An ICT-based intervention in the form of mobile apps seems to be quite popular. There currently exist 70,000 apps targeting people with various medical conditions (eg, an app to manage asthma symptoms) and health goals (eg, fitness and nutrition apps) [31]. Even though the importance of current diet mobile apps (eg, apps focusing on calorie intake) is hard to underestimate, it is plausible that apps related to CNE tailored to an adolescent population might play a role in the process of diet improvement.

Among all age groups, adolescents are typical early adopters of technology because of peer pressure, self-efficacy, and self-innovativeness [32]. Findings based on a nationally representative sample of adolescents indicate that the majority of adolescents use the internet to find health-related information [33]. Moreover, it has been shown that about 95% of adolescents in the United States, irrespective of gender, race, ethnicity, and socioeconomic background, either own or have access to a mobile phone [34]. Given potentially high interest, reach, a favorable adoption environment, convenience benefits related to time and location (eg, fitting cooking in the adolescents' busy schedules), and an increased call from the field of diabetes to include more tech solutions for providing nutrition care [35-37], a technology-enabled solution to assist adolescents with cooking would be highly desirable. There is, however, a dearth of evidence-based research that focuses on CNE-related mobile apps, particularly for adolescents [35].

A literature review of mobile apps for weight management by Azar et al found that out of 200 top ranked health and fitness apps (23 eligible apps) only 3 were focused on healthy cooking [38]. Furthermore, in our own search for commercially available apps for cooking on the iTunes App Store, we found about 100 relevant results almost all of which were related to cooking games. Most cooking and food related games tend to be centered around the entertainment value (eg, cooking or preparing food in virtual worlds) [39]. Even though the efficacy of such apps to teach cooking skills is yet to be evaluated, some preliminary studies did indicate that games may increase children's intake of fruits and vegetables [40]. Although the preliminary studies may be promising, unfortunately, it also seems that most of the commercially available apps do not address all components of CNE, particularly the knowledge about intake of fruits and vegetables, sodium, and added sugar, and are not tailored to the adolescent audience [41].

### Objective

Recent mobile health studies indicate a tendency to implement user-centered design (UCD) iterative technology development because of its deep focus on determining users' needs and the environment in which a technology might be used [42]. Mobile health interventions that are centered around a user or patient are highly responsive to users' preferences and, as a result, can induce higher level of engagement and smoother adoption [43,44]. Given the potential and nuances of the adolescent population, we root our effort in a UCD (and in some ways user-driven) approach to explore the design, functionalities, and potential of a technology-enabled solution to facilitate overweight and obese adolescents' self-care and provide them with CNE. The objectives of the study are to identify (1) user-centered functionalities (ie, a set of capabilities associated with a software [45]), (2) design requirements for a CNE technology (CNE-tech), and (3) contextual factors that could affect the adoption and use of technology.

Our study contributes to the existing literature by addressing the dearth of evidence-based research on CNE that leverages technology. We will also discuss our results in light of existing behavior change theories.

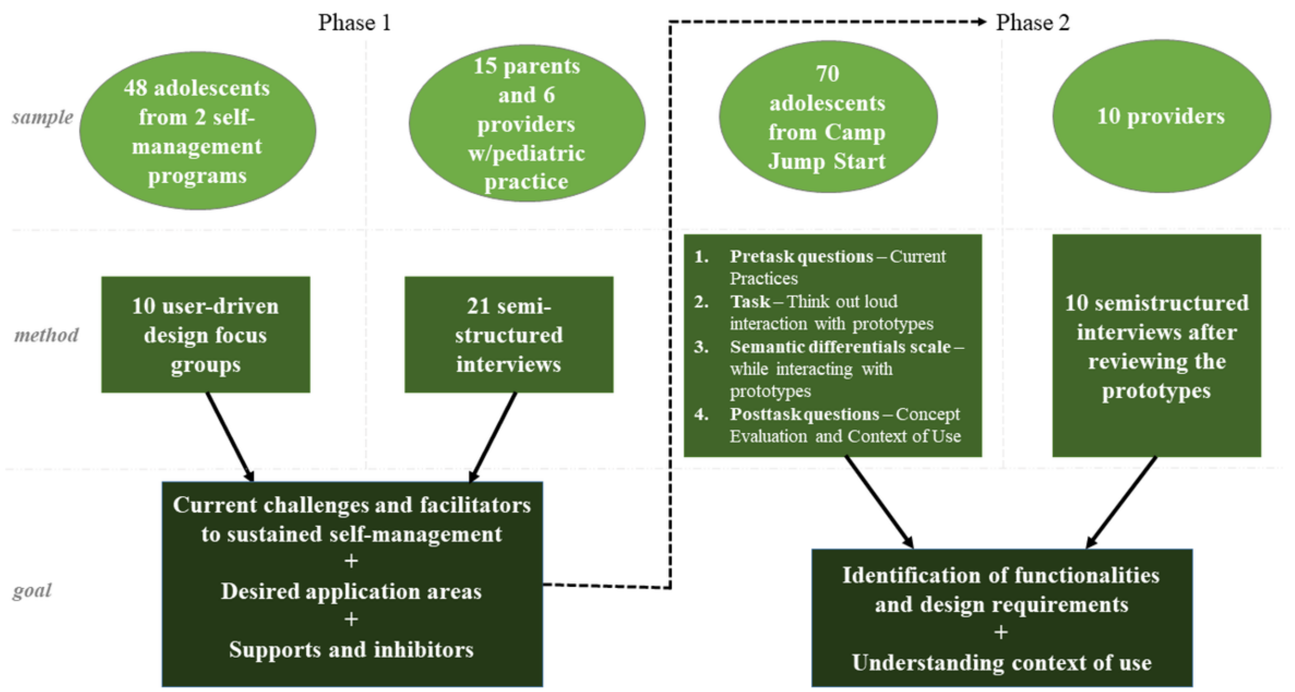
## Methods

### Study Design

In this design science study, we applied principles of UCD. We selected a UCD approach because it focuses on usability of the final technology product as demonstrated through learnability, efficiency, memorability, and satisfaction [46]. We identified multiple facets of usability of a mobile app–based intervention for overweight and obesity that aligned with the needs of our end users, adolescents. The study was guided (design and protocols) by constructs from the Unified Theory of Acceptance and Use of Technology (UTAUT)–performance expectancy (likelihood of meeting goals with Consumer Health Technology), effort expectancy (ease of use), and social influence (influence by others in the use of technology) [47].

Our user-centered study comprised 2 mixed-methods phases with phase 2 serving as the primary source of data for this paper. Figure 1 provides details on sample, method, and goal of each phase. The study protocol for both phases was approved by the Institutional Review Board of Saint Louis University. The adolescents were asked to assent before all data collection sessions. Reporting of qualitative data was guided by the Consolidated Criteria for Reporting Qualitative Research [48].

**Figure 1.** Multiphase research design.



### Phase 1

The goal of this phase was to detect desired app areas to assist with self-management as well as to identify design requirements and functionalities for such areas. It was also important to determine the context of self-management for the adolescents.

Focus groups in phase 1 focused on identifying possibilities for sustaining self-management, overcoming challenges and enhancing facilitators, conceptualizing desired app areas, and contextual supports and inhibitors to using technology. More detailed information about participants of the focus groups, criteria of selection, description of the protocol, and adolescents' initial feedback can be found in the study by Knoblock-Hahn et al [49]. Following focus groups with the adolescents, semistructured interviews with the parents and health care providers with pediatric practice took place and addressed the following topics: awareness of the adolescent's weight, barriers and facilitators for treatment of adolescent overweight and obesity, and perceived usefulness of and intent to use technology. More detailed information about semistructured interviews can be found in the study by Knoblock-Hahn et al [50]. As a result of the focus groups and semistructured interviews, 5 viable app areas were identified: (1) social networking, (2) motivation, (3) cooking (which the adolescents also referenced as *recipe builder*), (4) physical activity management, and (5) food management. Midfidelity wireframe screen design prototypes [51] were developed based on phase 1 data collection.

### Phase 2

Phase 2 was conducted to evaluate the adolescents' response to user-inspired prototypes of the app areas identified in phase 1 and detect more specific functionalities and design requirements pertaining to each of the 5 app areas. We recruited 70 adolescents from Camp Jump Start [52], a recognized, evidence-based adolescent weight loss summer camp program (led by a medical provider) [53]. Participants of the camp were aged between 9 and 18 years coming from 50 states and 23 countries. Some of the participants were from homeless and impoverished families, whereas others were coming from families with high socioeconomic status. Most of the participants were from the middle-class families.

Our study included a diverse group of black, white, and Hispanic adolescents from various socioeconomic backgrounds. Both males and females were included in the study. The age of the participants ranged from 12 to 17 years. The inclusion criteria were based on the age, computer use, and body mass index (participants were in 85$^{th}$-99$^{th}$ percentile range). The adolescents completed a series of activities covering from 2 to 4 of the 5 app areas identified in phase 1, as time allowed. We used a randomized *round-robin* rotation of the order of app areas introduced to each participant to assign app areas to 70 participants. This rotation was used to reduce any bias that may have resulted in the sequencing of viewing app areas and to facilitate coverage of all the apps with the aggregated pool of participants within the time constraints of each participant usability session. As a result of this rotation, 15 participants reviewed the CNE prototypes (n=15). For purposes of this paper, we also reviewed and included relevant data from the participants who reviewed other app areas for comments and references related to their current use of technology, cooking references, and general contexts of using technology to support their self-management.

### *Prototype Usability Testing*

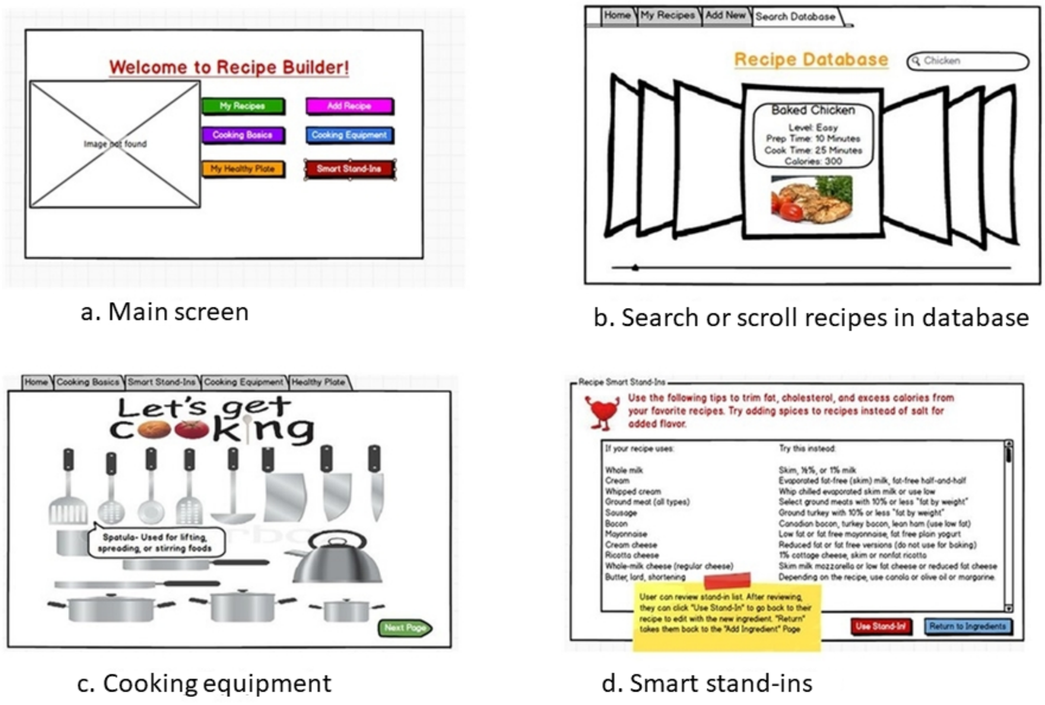
Participants who reviewed the CNE prototypes evaluated 5 wireframe screen design prototypes (low- to midfidelity) [51] for the cooking app. The prototypes were presented to them on laptop, tablet, and mobile phones (see Figure 2 for example screens). The adolescents were also provided with paper printouts of the screen designs.

The general sessions included the following series of interactions and tasks:

- **Preinteraction With Screen Mock-ups**
  (1) Sessions began with interview questions regarding the adolescents' use of technology, current cooking habits, and attitude toward using an app to assist with cooking.
- **During Interaction With Screen Mock-ups**
  (2) Think out loud review of the CNE-tech prototype.
  (3) Semantic differential scale [54] written assessment of the CNE prototypes.
  (4) Discussion of semantic differential scale responses with invitation to use markers, stickers, and other drawing supplies to identify issues and make changes to paper printouts of the prototype.
- **Postinteraction With Screen Mock-ups**
  (5) Detailed conceptual evaluation of the app.
  - Semistructured interview questions regarding the detailed design of the app for cooking.
  - Semistructured interview questions regarding the functions and general concept of using the app for cooking.
- (6) Semistructured interview questions regarding the context of use.

The usability session protocol further detailing each of the above activities is presented in Multimedia Appendix 1.

**Figure 2.** Culinary nutrition education-tech prototype screens.




### *Interviews With Health Care Providers*

During phase 2, we also conducted 10 semistructured, confidential telephone interviews (recorded audio tapes were destroyed after the study) with health care providers specialized in pediatric care and representing 4 states. The health care providers' practices included care of overweight or obese adolescents. The pediatric endocrinologist that served on the larger project provided an initial convenience sample of health care providers for recruitment for the interviews. In addition, we expanded our interview pool through the use of snowball sampling [55].

The health care providers spoke from their personal experience with treating obese and overweight adolescents and their knowledge of health behavior programs and resources that aimed to assist patients facing this health issue at large. Physicians' consent was received before arranging our interview sessions that lasted between 45 and 60 min. The health care providers received an electronic copy of the wireframed prototypes to reference during the call and the interviewer walked them through the app flow. The interview protocol focused on (1) their general reactions to the screen prototypes (the health care providers received the prototypes in advance), (2) challenges related to patient utilization of technology, (3) reactions to the prototypes, (4) their willingness to engage with technology, (5) ideas about the use and the integration of technology in summative patient reports in practice, and (6) opinions on the potential use of avatars and virtual agents. These interviews had limited coverage of the detailed design of each of the various app areas (eg, we did not go through preferred fonts and icons) as the health care providers took a more holistic assessment and recognized that they were not in a primary user role. Therefore, our results recognize specific statements directly or closely related to the CNE prototype and concepts, particularly related to safety and health issues; as well, the health care provider responses are most predominantly reflected in the context of use section of our results in recognition of their role in continuity of care.

## Data Analysis

After all usability walk-throughs were audio taped, transcribed, and reviewed for transcription errors, we conducted data analysis of deidentified transcripts using DEDOOSE, a qualitative data analysis application for data codification, classification, and treatment. Guiding principles proposed by Lee and Baskerville were applied to develop insights from the collected data [56].

We used constant comparison analysis to examine qualitative data [57,58]. First, 2 members of the research team independently coded interview transcripts and supplementary content. Our team deductively used the interview guide to formulate our high-level a priori coding schema (ie, predetermined coding). Researchers met regularly during this process to iteratively discuss emerging subcodes under each question and refine coding categories [59]. Intercoder disagreements regarding appropriate codes were resolved by consensus resolution, using an external qualitative expert to act as an auditor to make final determinations as needed. During the process, a few disagreements resulted from semantic issues of code name and precision, particularly whether or not to collapse detailed level coding into a higher code grouping. Consensus was ultimately reached among the 3 reviewers. We then carried out axial coding [59] to deductively collapse initial coding categories into specific functionalities and design requirements and contextual factors affecting use. Embedded in our interviewing and coding procedures, the validity and reliability of study data and interpretation were assessed following Lincoln and Guba's criteria for evaluating interpretive research [60,61].

## Results

### Overview

Of the 15 adolescents who specifically participated in usability testing for the CNE-tech and responded to demographic, closed-ended questions preceding interaction with the prototypes, all 15 stated that they have access to a desk or laptop computer and a mobile phone, and 14 (14/15, 93%) have access to a tablet. Regarding the use of technology, respondents indicated that they used laptop and desktop computers primarily for taking notes, Web surfing, social networking, games, emails, and school work. Mobile phones were used for fitness apps, games, music, communication, social networking, and Web searching. Finally, the adolescents indicated that they used tablets mainly for social networking, games, and music.

Regarding cooking experiences, 11 (11/15, 73%) adolescents indicated cooking *often* or *sometimes* whereas only 4 (4/15, 27%) indicated *never cooking*. A total of 13 (13/15, 87%) adolescents indicated cooking with recipes. Of the adolescents, 6 (6/15, 40%) used recipes found by their family members, 3 (3/15, 20%) used recipes from books, 3 (3/15, 20%) found recipes on the Web, and 9 (9/15, 60%) stated that they created their own recipes. In response to open-ended questions about their cooking experiences, the adolescents brought up lack of knowledge about nutritional content, lack of time for cooking, challenges with remembering steps for meal preparation, and problems with using appliances. Following are the results of usability testing.

Figure 3 overviews our results from participant interaction with the prototypes (think out loud, semantic differential scale, and conceptual design interview questions to explore the details of functionalities and design) and subsequent discussion on situations of use to saturate our discussion on (1) CNE functionalities, (2) CNE design requirements, and (3) context of use.

Figure 4 summarizes the results from the semantic differential scale that the adolescents completed while interacting with prototypes. The figure displays the marked semantic differential scale for the adolescents showing the mean position on the scale. Overall, the aggregated results of this scale indicated that the high-level prototypes were positively assessed by the participants in terms of various constructs of the UTAUT model: the adolescents' mean responses weighted toward useful, easy to understand, easy to learn, and aesthetic preferences, as demonstrated by the mean score on the semantic differential scale results from Figure 4. The results of this assessment also point to the relative attainment of the balance of *somewhere in between* childlike and adult design when it came to the look of the app. We extended the discussion of this balance into such areas as exploring appropriate icons and fonts to add to the evolving prototype, which we will discuss in the forthcoming design section.

**Figure 3.** Summary of findings.

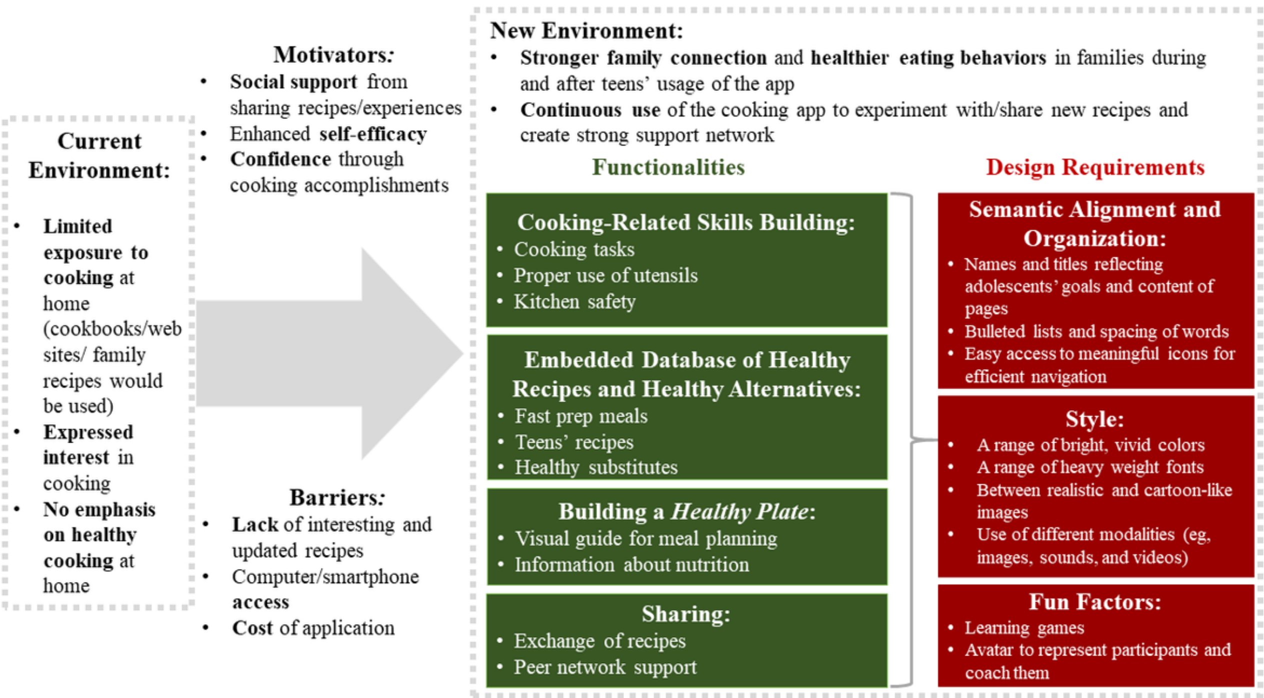
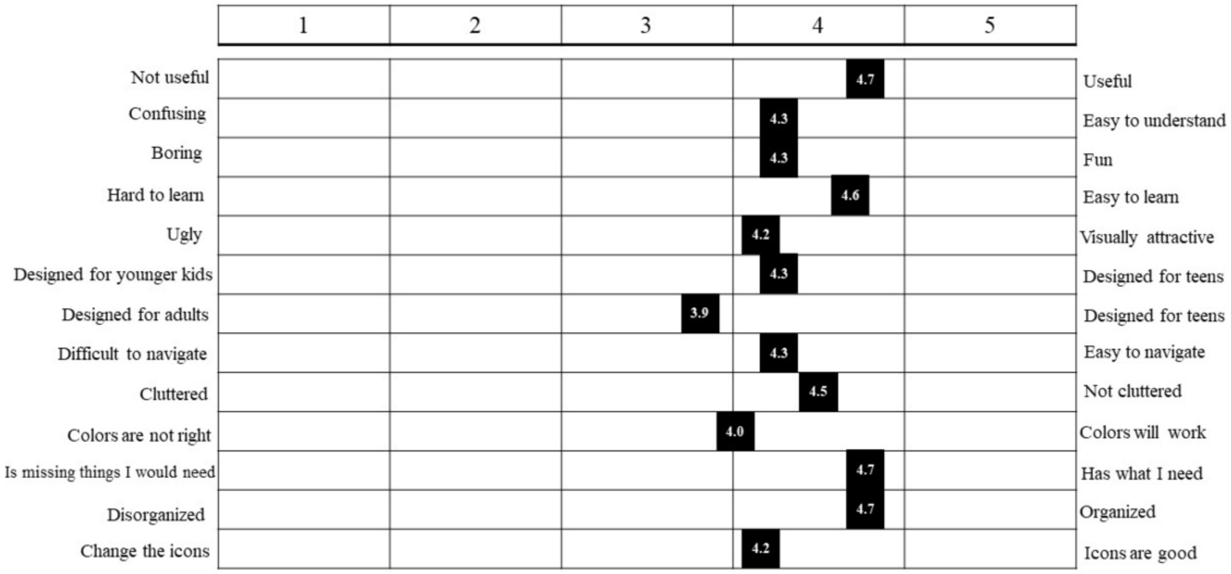


**Figure 4.** Semantic differential scale: the mean positions of adolescents.




## Key Functionalities

The adolescents indicated a series of functions that could be useful to facilitate their self-management as shown below.

### *Cooking-Related Skills Building*

Many adolescents pointed out that they wanted to attain basic kitchen knowledge through the use of CNE-tech. They wanted to know *what knives to use, like what utensils you should use* and *what temperature to put the oven on*. Regarding specific cooking skills, they wanted to learn not just how to prepare the food but also creative ways to present the food. As 1 adolescent stated:

> *I would want to learn cool ways to prepare fruit like have you ever seen those design fruits where you put one fruit in the other fruit like it is some kind of a pattern. I think of creative ways to prepare food.*

Moreover, the health care providers brought up the importance of kitchen safety:

> *People need to know how to defrost frozen foods, safely...They need to know how to use a knife safely. You need a little session on some of those in particular with the kids.*

### *Embedded Database of Healthy Recipes and Healthy Alternatives*

The adolescents stated that the CNE-tech should contain a stocked database of predetermined healthy recipes. They also indicated that the database should include recipes for all meal times and snacks and that every recipe should include estimated preparation times (ideally short) because of their tight schedules by stating “the recipes need to not take a lot of time, thirty minutes or less, if possible. You know, it needs to involve low-cost ingredients.” They further indicated that recipe metainformation should include complexity level and nutritional information. Furthermore, the adolescents indicated that they wanted pictures to accompany step-by-step instructions.

In addition to prepopulated recipes, adolescents also wanted to be able to add their own recipes to the database. In making these additions or changes, adolescents wanted healthy guidance, namely, that the app would suggest substitutions for unhealthy ingredients when building or reviewing a recipe. As one of them stated:

> *Because I do alter up a lot of things when I’m cooking and substitute it for different things. Also, it’ll be helpful to have the healthier substitutions thing going so that way I can kind of be like, “Yay healthy.”*

This preference for guidance extended to some adolescents suggesting alerts in case of an *unhealthy choice*. Furthermore, the participants suggested functionality that would convert recipes they found outside of the app to healthy ones leveraging these substitutes. As stated by one adolescent:

> *I would absolutely love it if there was an app, that you could plug in your recipes, you know, even just a normal recipe that you haven’t modified to make healthy, you just want to know.*

One participant suggested that the CNE-tech could have links to specific external websites that would allow them to pick and convert recipes into healthier alternatives.

Regardless of the original source of the recipe, many adolescents indicated the desire to add personal notes and pictures of meals they prepared to the recipes in their database.

### *Building a Healthy Plate and Learning About Nutrition*

Given that the adolescents were already in a program making their way toward behavior change, it was not surprising that they understood the concept of a healthy meal and attention to daily food consumption. To help themselves aim for a healthy meal in selecting recipes, the adolescents described a plate metaphor, a proportional puzzle, that the app would consider aggregated food choices to help them balance their meals. A proportional plate function was referenced as a visual guide for meal planning that would indicate the missing nutritional elements and provide suggestions to complete building a healthy plate.

The adolescents also indicated that the CNE-tech could provide a means to learn more about nutrition:

> *[We can learn] the nutrition about the little star next to things that basically says it's a 2000 calorie diet, when you know most people were on a 1500 calorie diet.*

### *Sharing*

The adolescents mentioned that they wanted the CNE to have a private peer network that could enable sharing and posting pictures or information about recipes and their experiences with family and friends and camp counselors. One of the participants said:

> *I don't normally share recipes on Facebook, but with this site, I probably might because they are going to be healthy and then I can show my friends "Oh, look I am eating healthy you should to."*

The adolescents indicated that this peer sharing of cooking and eating accomplishments would motivate them to eat and cook healthier.

## Key Design Requirements

The adolescents expressed a series of preferences in technology design particularly related to alignment between names and content, page layouts, colors, fonts, image styles, use of modalities, games, and avatars.

### *Semantic Alignment and Organization*

The adolescents were sensitive to semantic alignment of the app name and headers. They preferred that the name of the app reflect the goal of the app (ie, *Health Life, Health Track*, and *Nutrition Buddy*) and titles of individual pages match the content of such pages.

Furthermore, the participants indicated that the content within pages should consider the use of bulleted lists, white space and other organizing structures to present information clearly. A tabular structure for organizing recipes was mentioned:

> *A tab for...recipes because there would be a lot of them right, then you could have my friend's recipes as another tab.*

As for the overall layout of the app, the adolescents wanted easy access to organized and meaningful icons to facilitate efficient navigation. As one of them stated:

> *When you open it, it can have your different choices at the bottom of a screen, and one of them could be that icon for restaurant, and then another one for tips or something, and then another one for the game, and then another one for whatever else we want.*

### *Style*

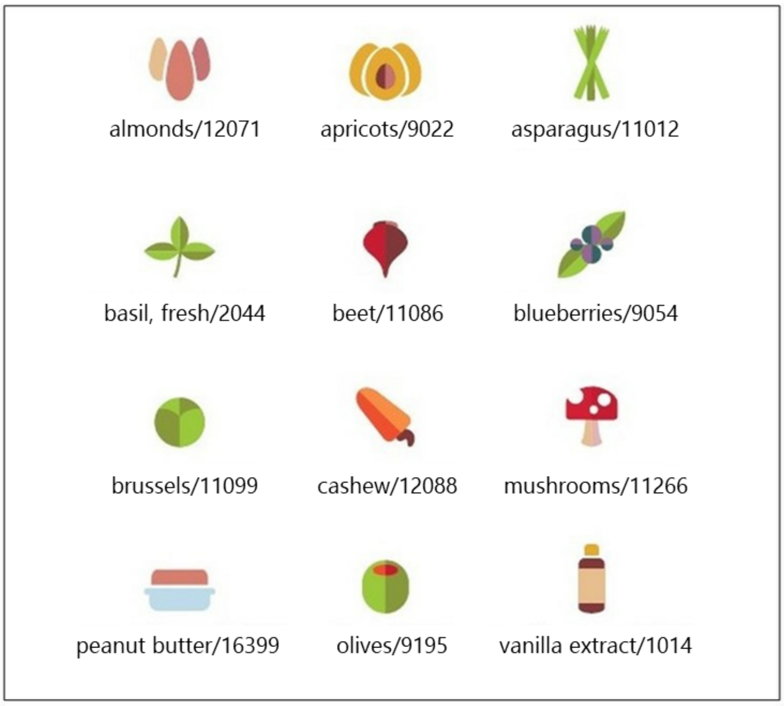
The adolescents expressed enthusiasm about having a range of bright, vivid colors throughout the app. They also indicated that they wanted to choose among different fonts and were particularly interested in font types, such as Comic Sans, Custom Bubble, Hawaiian Punk, Disco, Island of Misfit Toys, Lucida Handwriting, Montgomery, and Princetown LET, Star Guide; research studies indicate that these fonts (which fall into the script or funny category established in past studies) convey a happy and creative emotional message [62]. Figure 5 shows examples of contrasting font preferences. The participants also noted that the fonts chosen should be consistent throughout the app to unify it.

The adolescent responses to image styles indicated that a careful balance between realism and playful was needed—semireal cartoon. Figure 6 contains examples of the final design for the images to strike a balance between adult and childlike based on the adolescents' responses from a semantic differential scale described previously and their responses to various style image options. To present these images, some participants indicated a preference of borders around the images and further specified a preference for circular (rounded corner) rather than rectangular shaped borders within the CNE-tech.

**Figure 5.** Examples of fonts.

| Preferred | Nonpreferred |
|---|---|
| Lucida Handwriting | Jazz LET |
| Island of Misfit Toys | Bauhaus |
| Hawaiian Punk | Blair MDITC |
| Mistral | Bangla MN |

**Figure 6.** Example of the final design of the food images.




The participants also expressed their interest in the use of different modalities, such as images, sounds, and videos, within the app. They indicated that these modalities could complement and enliven various app areas, for example, interactive video clips could accompany challenging recipes. As for sounds, the adolescents suggested sound effects that resemble real sounds related to cooking, as 1 adolescent mentioned:

> *When something is in the oven, it will be, like sizzle too, but louder ones. And when you're washing your hands, shhh, like that sound. And when you're cutting (pounding on the table) sound.*

### *Fun Factors: Gaming and Avatars*

The adolescents also revealed their interest in the gaming component of the app. Learning games (ie, identifying unhealthy food options) were brought by the participants, meaning that they wanted the gaming component to tie to the learning objective of the app, cooking literacy. One of the participants stated:

> *I think that it defeats the purpose of a website that helps kids, the games would be just something they would play for fun, it wouldn't be really helping their health or unless they were like inspired by the game, but I don't see that happening.*

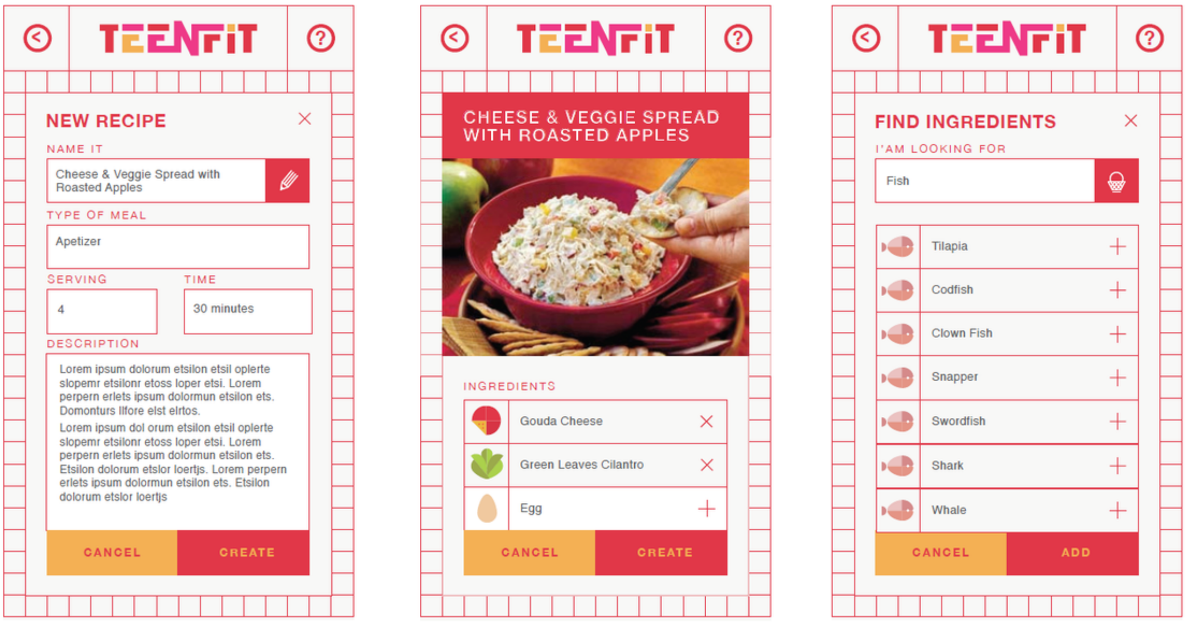
Finally, the adolescents indicated that avatars could serve multiple purposes in the app (eg, animated representation of the participants or a coach demonstrating cooking tasks). The participants indicated that avatars could bring some fun factors into the app; as one of them said:

> *You could favorite one of your recipes and then when you got close to one, it [Avatar] would jump [up and] down silently.*

## Evolved Prototypes

Figure 7 provides examples of the evolved screen design for a mobile phone platform based on the design sessions with the adolescents.

**Figure 7.** Modified screen designs.




## Context of Use

The CNE-tech was of high interest to all parties involved in this study. The health care providers expressed particular enthusiasm, as one of them stated,

> *It never really crossed my mind because, of course, my boys like to cook and they wanted to learn from us, and then we are really cooking so, that actually never really crossed my mind that these kids when they leave home don't know how to cook, so they will be more likely [to eat] fast foods and stuff like that.*

Furthermore, the parents engaged in phase 1 of the study thought that the CNE-tech that could help their adolescent was beneficial and fun:

> *Something to help your child with meal plans?* [Interviewer]
>
> *Yeah, I think they would definitely use that because they are very computer and internet savvy. They would consider that fun I think. They would have control over choices and decisions.* [Parent]

The adolescents were excited about cooking in general, as some of them already cooked using recipes from family, books, websites, and even food packages. They also revealed that sometimes they created their own recipes for food items, while acknowledging that these experiments were not always successful, as 1 participant said:

> *I've tried, but they're kind of a train wreck. They don't taste very good either, but I try though.*

When the adolescents compared this app with the existing cookbooks and other websites or apps, they pointed out some differences from what they used, such as having recipes in one place, calorie counting, and nutritional information that is simple and comprehensive. As 1 participant indicated:

> *So this is a lot like* ***Epicurious*** *but it is a little more easy to understand and it gives you substitutes if you do not want to use what it says.*

According to the adolescents and their parents, lifestyle management that involves tracking food intake and cooking ability would improve as a result of using the app. The adolescents indicated they would cook more with the CNE-tech. They also recognized that using the CNE-tech to enact healthy cooking and eating would be part of their self-management. One of the participants said:

> *I think it would work because then you can kind of see exactly what you're putting into your body. If your doctor says, "Oh, you need more calcium" or something, you would know that your food has calcium in it.*

Some of the parents, for example, were particularly interested in the healthy calorie intake per meal, as 1 of them brought up:

> *Yeah, what am I looking at here? How many grams of fat would be in this? How many grams of sugar? How many calories? You know, what would be a decent serving size of this to stay within a decent amount of calories?*

Furthermore, the adolescents were particularly enthusiastic about the option of sharing their progress and recipes with their social network (specific mention included parents, siblings, grandmother, friends, camp counselors, and doctors). One of the adolescents said:

> *Well I'd like my friends and my family to go [into the app] and they can go in separately and alone, but I could go on with my mom and it might be that, "Hey. Do you want any of recipes" and then she can help me look at different stuff.*

Such aspects of social sharing of recipes were also brought up by the health care providers in the interviews:

> *I think that regarding the recipe, like if they create a recipe, or choose a family recipe or whatever that they adding to it, there will be a way that they could share that with their social group and all the social age kind of stuff.*

The participants were optimistic to use the CNE-tech every day around dinner time. One of the adolescents said:

> *I'd have time to use it, probably, I'd try to use it every night just to see if I could make something or create something.*

Furthermore, some adolescents stated that they would generally use healthy behavior technology multiple times a day during the first month and keep using it afterward. The parents were supportive of such an idea and expressed willingness to work with their children on recipes. One of the parents stated:

> *Her having the app so that I can say, "Hey, find something that's approved on the app, yeah, well go have it let's do it." So, she has resources to find things that she likes, or recipes or ways to cook her favorite foods that she can give to me.*

The adolescents indicated that their usage, however, might decline because of the lack of interesting and updated recipes or lack of information on healthy alternatives and substitutions. As one of them stated:

> *But one of the times where I might be less interested in the website would be if the recipes were getting less interesting or if it is starting to become stuff that I didn't like.*

In addition, potential barriers, such as access to technology or cost of technology could affect usage.

## Discussion

### Behavioral Change Theory Alignment

Eating home-cooked meals along with the recommended amounts of fruits and vegetables is associated with an increase in culinary skills and healthy cooking behaviors that could combat obesity and overweight [5]. Our results demonstrate the functional needs and design preferences from stakeholders (eg, adolescents, parents, and health care providers) for a CNE-tech that focuses on sustainable nutrition planning while capturing the fun, rewarding side of cooking. We then identified contextual factors, such as strong support system, social network, and motivators to increase self-esteem and self-efficacy, that could affect the participants' use of CNE-tech and sustain usage thereof over time.

Understanding requirements for the design and functionalities of the CNE-tech provides us with certain insights into the mental models of our target users, so we have a better chance of creating a technology-enabled solution that would appeal to our audience. In addition to practical implications, our results demonstrated some alignment with the existing behavioral change theories, meaning that identified requirements exhibited high potentiality to induce changes in the participants' health behaviors.

Our functional and design guidelines echo various aspects of behavioral change theories. Reflection of various constructs from behavioral change theories in mobile app interventions is not a novel idea, as it was demonstrated in the recent review on commercially available cooking apps in the Apple store [38]. Highly ranked diet tracking apps emphasize behavioral strategies, such as perceived social norms, self-monitoring, and realistic goal setting [63]. In our study, functionalities and design requirements are also tied to similar constructs on which behavioral strategies might be grounded.

The adolescents' preferences toward various elements of design requirements and functionalities appear to align with constructs of the Social Cognitive Theory (SCT), one of the main behavior change theories focusing on the ongoing change in a person's behavior [64], and the Fogg Behavior Model (FBM), which explores the effect of persuasive design on people's behaviors [65].

In the SCT, the concept of reciprocal causation implies that there exists an ongoing, impactful interaction among environment, person, and behavior [64]. Our findings have shown that the adolescents would be eager to become more responsible for family meal preparation as they build up their food literacy using our CNE-tech. Exposing their families to dietary changes, the adolescents may alter their environment by effecting their relationships with their families along with healthy habits of their families. As the parents become more involved in dietary changes of their children (eg, buying ingredients, assisting the adolescents with certain cooking tasks, or participating in social and support networks), they will spend more time with the adolescents, build stronger family connections over time, and adopt healthy eating behaviors.

Before they induce any change in their environment and their own lifestyle, the adolescents need to have necessary skills to do so. Their behavior capability, referring to skills mastery in the SCT [64], would be enriched by cooking-related skill building (ie, learning about basic utensils or creating their own recipes). Design of pages and embedded database of healthy recipes in the CNE-tech prototype align with the concepts of ability from the FBM [65]. According to Fogg, ability in the persuasive technology implies the power of simplicity that makes behavior easier to do. This concept was reflected in highly organized, step-by-step recipes and easy-to-use embedded database of healthy recipes desired by the adolescents.

Trigger, another component of the FBM, is a vital element of persuasive technology that functions to activate a person's behavior by, for example, inspiring hope or highlighting fear [65]. This construct was reflected in the adolescents' requirements for the naming of app and page titles that should evoke the adolescents' vision of their future self.

Design requirements also show the desire for the extensive use of multimedia, such as pictures, audios, and videos, and such requirements align with the concept of observational learning from the SCT [64], meaning that adolescents are eager to learn by watching and listening to others in the lessons. Different elements of the interface, either bright colors in images or realistic sounds in videos, should create some kind of a happy, fun, and energetic feeling, yet they should not be too childish for adolescents. All combined, these elements should contribute to forming and maintaining adolescents' weight goals and expectations, another important component of the SCT [64].

One of the major constructs in the SCT is self-efficacy, a cognitively based source of motivation that encourages people to persist in their efforts [66]. After accomplishing a given level of performance, adolescents may no longer be satisfied with their progress level and take further steps toward higher attainments. This aspect of building self-efficacy was indicated by their general interest to use the CNE-tech and the gamification part, which would allow the adolescents to set incremental goals and build their self-efficacy in a fun, yet challenging way. Gamification along with using an avatar can also induce internal reinforcement [64] of adolescents' behavior as they reward themselves by building up their cooking confidence and undertaking more challenging recipes.

External reinforcement [64] of adolescents' behaviors stems from their participation in social and support networks. As adolescents share their experiences with and receive feedback from their family and friends, they might be exposed to social persuasion, which can contribute to the successful outcomes achieved through corrective performance, in the form of verbal encouragement [66]. In the FBM, Fogg discusses the idea of motivation in the form of social acceptance [65], which seems to align with adolescents' use of social and support networks where they share their experiences and may feel more socially accepted by their peers and families as they continue their journey toward building healthy habits.

### Practical Implications

Our results bring to light aspects of functional and design misalignment that some existing cooking and weight management technologies have with overweight adolescents' needs and preferences for maintaining healthy eating. For example, some existing apps record users' data, but focus primarily on the number of consumed calories rather than nutritional value of foods. Our data indicate that adolescents need and want to know more about food and cooking than calories. A cooking app with extended information can introduce the adolescents to nutrient-dense foods. Adolescents can use CNE-tech to not only cook at home and at the home of someone in their network but also guide their food choices in other settings. Benefits can be gained from away-from-home meals, such as packing lunches for school or selecting healthier food choices in restaurants or school cafeterias when options are available. Moreover, some programs on the market focus on selling branded, prepared meals. Such programs could provide customers with a healthy balance, yet failed to teach food literacy, which was clearly supported by the study participants. It seems that many available cooking apps expect users to have at least some experience in cooking, limiting their applicability to adolescents, who seem to be novices in many cases. Therefore, adolescents want the app to start from the basics (eg, cutting, frying, using utensils, and understanding oven temperatures) and help them improve their cooking skills over time. Current cooking apps, however, seem to overall lack such functions and the preferred design, creating the need for tools that embody the results of this study.

### Limitations

Our study is subject to various limitations affecting generalizability. Identified functionalities and design preferences could be more relevant to the segment of adolescent population that participated in our study, rather than to the overall population of overweight adolescents. For both phases of this study, our focus was on looking at healthy behavior technologies to extend punctuated or short-lived programs. This means that adolescents who participated in the study represented a population of adolescents enrolled in healthy behavior programs, meaning that they and their parents were aware of the weight issue and understood the need for behavioral modification. These adolescents were actively seeking change, and readers should recognize that our study might again not be generalizable to other overweight or obese adolescents in the country, who may not yet be receptive to or actively seeking change. Furthermore, the study is based on the assumption that adolescents are already using technologies and will be doing so in the future. Such assumption might not hold in some contexts, as some families might be more restrictive about their children using technologies or the capacity of the technologies the adolescents have access to might be incompatible with the CNE-tech for some reason.

Another assumption that might not hold is the access to healthy food options. CNE-tech might be challenging to use in communities that have limited access to healthy and nutritious food because of financial and other constraints. There, however, exist public health programs, such as the Healthy Food Financing Initiative and Supplemental Nutrition Assistance Program [67], that aim to provide the underserved communities with access to healthy and affordable food options.

It is also of note that to extend further reach of these technologies and programs to those in lower socioeconomic status, iterations of an app informed by the larger study are under development. The effort is specifically targeted to extend the reach of the evidence-based healthy programs to adolescents who are not able to attend these programs because of various limitations and socioeconomic constraints. The concept will be tested in late 2019 and will couple functionalities of the technology with virtual coaching and guidance from professionals involved in the on-site evidence-based healthy behavior programs.

Future research may aim to address the above-mentioned limitations and the role of national public health initiatives to support generalizability of our results.

### Conclusions

As adolescent overweight and obesity collectively constitute one of the major health issues in our population [1], we need to understand the importance of behavioral health interventions that could cause a sustainable change in self-care. The overall decline of culinary skills and home cooking in the US population stands in the way of combating obesity and overweight and associated health conditions. Our study leveraged a multistakeholder (eg, adolescent, parent, and provider) user-centered approach to tailor CNE-tech requirements and design to adolescents' preferences and needs and also identified contextual factors affecting the potential use. Overall, our study indicates that adolescents, health care providers, and parents see that CNE-tech solutions that address the features they specified have the potential to facilitate the self-management of overweight and obesity. The culinary skills acquired may be additive or exponential as upon achieving a critical level of food literacy, an overweight adolescent can not only prepare food for themselves but also share healthy meals, teach others to cook, and actively advocate for their own healthy nutrition needs. Furthermore, the acquisition of food literacy and cooking confidence might have some carryover effects to other forms of self-management and self-esteem.

## Acknowledgments

The study was funded by Saint Louis University's Center for Sustainability grant. The authors greatly appreciate the work of Jean Huelsing, Camp Jump Start (Living Well Foundation) founder, for her cooperation and insight supporting this study and Dr Toree Malasanos (MD–pediatric endocrinologist) for her consultation and insight supporting this study.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

User centered design session protocol.

[PDF File (Adobe PDF File), 196KB - jmir_v21i7e14430_app1.pdf ]

## References

1. Carroll MD, Navaneelan T, Bryan S, Ogden CL. Centers for Disease Control and Prevention. 2015 Aug. Prevalence of Obesity Among Children and Adolescents in the United States and Canada URL: http://www.cdc.gov/nchs/products/databriefs/db211.htm [WebCite Cache ID 753fv01nq]
2. Al-Goblan AS, Al-Alfi MA, Khan MZ. Mechanism linking diabetes mellitus and obesity. Diabetes Metab Syndr Obes 2014;7:587-591 [FREE Full text] [doi: 10.2147/DMSO.S67400] [Medline: 25506234]
3. You J, Choo J. Adolescent overweight and obesity: links to socioeconomic status and fruit and vegetable intakes. Int J Environ Res Public Health 2016 Mar 9;13(3):pii: E307 [FREE Full text] [doi: 10.3390/ijerph13030307] [Medline: 27005654]
4. Hruby A, Hu FB. The epidemiology of obesity: a big picture. Pharmacoeconomics 2015 Jul;33(7):673-689 [FREE Full text] [doi: 10.1007/s40273-014-0243-x] [Medline: 25471927]
5. Güngör NK. Overweight and obesity in children and adolescents. J Clin Res Pediatr Endocrinol 2014 Sep;6(3):129-143 [FREE Full text] [doi: 10.4274/Jcrpe.1471] [Medline: 25241606]
6. Sahoo K, Sahoo B, Choudhury AK, Sofi NY, Kumar R, Bhadoria AS. Childhood obesity: causes and consequences. J Family Med Prim Care 2015;4(2):187-192 [FREE Full text] [doi: 10.4103/2249-4863.154628] [Medline: 25949965]
7. Anderson PM, Butcher KE. Childhood obesity: trends and potential causes. Future Child 2006;16(1):19-45. [doi: 10.1353/foc.2006.0001] [Medline: 16532657]
8. Utter J, Larson N, Laska MN, Winkler M, Neumark-Sztainer D. Self-perceived cooking skills in emerging adulthood predict better dietary behaviors and intake 10 years later: a longitudinal study. J Nutr Educ Behav 2018 May;50(5):494-500 [FREE Full text] [doi: 10.1016/j.jneb.2018.01.021] [Medline: 29525525]
9. Wolfson JA, Frattaroli S, Bleich SN, Smith KC, Teret SP. Perspectives on learning to cook and public support for cooking education policies in the United States: a mixed methods study. Appetite 2017 Dec 1;108:226-237. [doi: 10.1016/j.appet.2016.10.004] [Medline: 27720707]
10. Okrent AM, Kumcu A. Economic Research Service. 2016. US Households' Demand for Convenience Foods URL: https://www.ers.usda.gov/webdocs/publications/80654/err-211.pdf?v=0 [WebCite Cache ID 753pPxtEX]
11. Siddiqua AA, Alem MS. Impact of situational factors in students' preference of fast food - an empirical study. Int J Res Com Manag 2018 Jul 1;9(7):9-11 [FREE Full text]
12. Contini C, Boncinelli F, Gerini F, Scozzafava G, Casini L. Investigating the role of personal and context-related factors in convenience foods consumption. Appetite 2018 Dec 1;126:26-35. [doi: 10.1016/j.appet.2018.02.031] [Medline: 29571961]
13. Saghaian S, Mohammadi H. Factors affecting frequency of fast food consumption. J Food Dist Res 2018 Mar;49(1):22-29 [FREE Full text]
14. Janssen HG, Davies IG, Richardson LD, Stevenson L. Determinants of takeaway and fast food consumption: a narrative review. Nutr Res Rev 2018 Dec;31(1):16-34. [doi: 10.1017/S0954422417000178] [Medline: 29037273]
15. Region of Waterloo. 2015. Food Skills of Waterloo Region – Changes Over 6 Years URL: https://www.regionofwaterloo.ca/en/health-and-wellness/resources/Documents/FoodSkills_WR.pdf [WebCite Cache ID 753weJqCx]

16. Raber M, Chandra J, Upadhyaya M, Schick V, Strong LL, Durand C, et al. An evidence-based conceptual framework of healthy cooking. Prev Med Rep 2016 Dec;4:23-28 [FREE Full text] [doi: 10.1016/j.pmedr.2016.05.004] [Medline: 27413657]
17. Perera T, Frei S, Frei B, Wong SS, Gerd B. Improving nutrition education in US elementary schools: challenges and opportunities. J Educ Pract 2015;6(30):41-50 [FREE Full text]
18. Snyder SA, Dougall AL. Assessment of nutrition topics for education in college-aged adults. J Food Nutr Res 2018 Jan 6;6(8):525-530. [doi: 10.12691/jfnr-6-8-7]
19. Kris-Etherton PM, Akabas SR, Bales CW, Bistrian B, Braun L, Edwards MS, et al. The need to advance nutrition education in the training of health care professionals and recommended research to evaluate implementation and effectiveness. Am J Clin Nutr 2014 May;99(5 Suppl):1153S-1166S [FREE Full text] [doi: 10.3945/ajcn.113.073502] [Medline: 24717343]
20. Guthrie JF. Integrating behavioral economics into nutrition education research and practice. J Nutr Educ Behav 2017 Dec;49(8):700-5.e1. [doi: 10.1016/j.jneb.2016.09.006] [Medline: 27756594]
21. McMullen J, Ickes M. The influence of a campus-based culinary, nutrition education program, 'College CHEF,' on college students' self-efficacy with cooking skills and nutrition behaviors. Build Healthy Acad Communities J 2017 Nov 27;1(2):61. [doi: 10.18061/bhac.v1i2.5987]
22. Kerrison DA, Condrasky MD, Sharp JL. Culinary nutrition education for undergraduate nutrition dietetics students. Brit Food J 2017 May 2;119(5):1045-1051. [doi: 10.1108/BFJ-09-2016-0437]
23. Abraham S, Noriega B, Shin JY. College students eating habits and knowledge of nutritional requirements. J Nutr Hum Health 2018;2(1):13-17 [FREE Full text]
24. Thomas HM, Irwin JD. Cook it up! A community-based cooking program for at-risk youth: overview of a food literacy intervention. BMC Res Notes 2011 Nov 15;4:495 [FREE Full text] [doi: 10.1186/1756-0500-4-495] [Medline: 22085523]
25. Wrieden WL, Anderson AS, Longbottom PJ, Valentine K, Stead M, Caraher M, et al. The impact of a community-based food skills intervention on cooking confidence, food preparation methods and dietary choices - an exploratory trial. Public Health Nutr 2007 Feb;10(2):203-211. [doi: 10.1017/S1368980007246658] [Medline: 17261231]
26. Bernardo GL, Jomori MM, Fernandes AC, Colussi CF, Condrasky MD, Proença PR. Positive impact of a cooking skills intervention among Brazilian university students: six months follow-up of a randomized controlled trial. Appetite 2018 Dec 1;130:247-255. [doi: 10.1016/j.appet.2018.08.014] [Medline: 30118784]
27. Utter J, Denny S, Lucassen M, Dyson B. Adolescent cooking abilities and behaviors: associations with nutrition and emotional well-being. J Nutr Educ Behav 2016 Jan;48(1):35-41.e1. [doi: 10.1016/j.jneb.2015.08.016] [Medline: 26411900]
28. McMullen J, Ickes M, Noland M, Erwin H, Helme D. Development of 'College CHEF,' a campus-based culinary nutrition program. Am J Health Educ 2016 Dec 29;48(1):22-31. [doi: 10.1080/19325037.2016.1250016]
29. do Amaral EM, de Carvalho SV, dos Santos CC, Toral N. Nutritional interventions for adolescents using information and communication technologies (ICTs): a systematic review. PLoS One 2017;12(9):e0184509 [FREE Full text] [doi: 10.1371/journal.pone.0184509] [Medline: 28961248]
30. 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]
31. Research 2 Guidance. 2015. mHealth App Developer Economics 2015 URL: https://research2guidance.com/product/mhealth-developer-economics-2015/
32. Lee SY. Examining the factors that influence early adopters' smartphone adoption: the case of college students. Telemat Inform 2014 May;31(2):308-318. [doi: 10.1016/j.tele.2013.06.001]
33. Wartella EA, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella AR. Teens, health and technology: a national survey. Media Commun 2016 Jun 16;4(3):13-23. [doi: 10.17645/mac.v4i3.515]
34. Anderson M, Jiang J. Pew Research Center. 2018. Teens, Social Media & Technology 2018 URL: https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [WebCite Cache ID 78j9Noh83]
35. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
36. Stein K. Remote nutrition counseling: considerations in a new channel for client communication. J Acad Nutr Diet 2015 Oct;115(10):1561-1576. [doi: 10.1016/j.jand.2015.06.375] [Medline: 26271692]
37. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. J Hum Nutr Diet 2017 Dec;30(4):439-452. [doi: 10.1111/jhn.12446] [Medline: 28116773]
38. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
39. Wei J, Cheok AD, Martinez XR, Tache R, Zhu Q. Foodie: Play With Your Food Extend Social Cooking Game With Novel Edible Interface. In: Proceedings of the International Games Innovation Conference. 2011 Presented at: IEEE'11; November 2-3, 2011; Orange, California p. 59-61. [doi: 10.1109/IGIC.2011.6115132]

40. Thompson D, Bhatt R, Vazquez I, Cullen KW, Baranowski J, Baranowski T, et al. Creating action plans in a serious video game increases and maintains child fruit-vegetable intake: a randomized controlled trial. Int J Behav Nutr Phys Act 2015 Mar 18;12:39 [FREE Full text] [doi: 10.1186/s12966-015-0199-z] [Medline: 25890060]
41. Hingle M, Patrick H. There are thousands of apps for that: navigating mobile technology for nutrition education and behavior. J Nutr Educ Behav 2016 Mar;48(3):213-8.e1. [doi: 10.1016/j.jneb.2015.12.009] [Medline: 26965099]
42. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
43. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. J Diabetes Sci Technol 2013 Jul 1;7(4):1039-1056 [FREE Full text] [doi: 10.1177/193229681300700429] [Medline: 23911188]
44. 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]
45. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/functionality [accessed 2019-04-01] [WebCite Cache ID 774MZSyT1]
46. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016 Jun;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]
47. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
48. 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]
49. Knoblock-Hahn AL, Wray R, LeRouge CM. Perceptions of adolescents with overweight and obesity for the development of user-centered design self-management tools within the context of the chronic care model: a qualitative study. J Acad Nutr Diet 2016 Dec;116(6):957-967. [doi: 10.1016/j.jand.2015.08.022] [Medline: 26477975]
50. Knoblock-Hahn AL, LeRouge CM. A qualitative, exploratory study of predominantly female parental perceptions of consumer health technology use by their overweight and/or obese female adolescent participating in a fee-based 4-week weight-management intervention. J Acad Nutr Diet 2014 Apr;114(4):570-577. [doi: 10.1016/j.jand.2013.11.021] [Medline: 24507626]
51. Engelberg D, Seffah A. A framework for rapid mid-fidelity prototyping of web sites. In: Hammond J, Gross T, Wesson J, editors. Usability: Gaining a Competitive Edge. Boston, MA: Springer; 2002:203-216.
52. Camp Jump Start: Weight Loss Camp for Teens - More than a Fat Camp. URL: https://www.campjumpstart.com/ [accessed 2018-02-01] [WebCite Cache ID 756jCCQgt]
53. Huelsing J, Kanafani N, Mao J, White NH. Camp jump start: effects of a residential summer weight-loss camp for older children and adolescents. Pediatrics 2010 Apr;125(4):e884-e890. [doi: 10.1542/peds.2009-1007] [Medline: 20194273]
54. Frey B. The Sage Encyclopedia of Educational Research, Measurement, and Evaluation. Volume 1-4. Thousand Oaks, CA: Sage Publications; 2018.
55. Berg S. Snowball sampling. In: Encyclopedia of Environmetrics. Chichester: Wiley; 2002.
56. Lee AS, Baskerville RL. Generalizing generalizability in information systems research. Inform Syst Res 2003 Sep;14(3):221-243. [doi: 10.1287/isre.14.3.221.16560]
57. Glaser BG. The constant comparative method of qualitative analysis. Soc Probl 1965 Apr;12(4):436-445. [doi: 10.2307/798843]
58. 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]
59. Saldana J. The Coding Manual for Qualitative Researchers. Second Edition. Los Angeles, CA: Sage Publications; 2013.
60. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. Educ Technol Res Dev 1981;29(2):75. [doi: 10.1007/BF02766777]
61. Guba EG, Lincoln YS. Naturalistic Inquiry. First Edition. Volume 75. Newbury Park, CA: Sage Publications; 1985.
62. Shaikh D, Chaparro B. Digital font and reading. In: Series on Language Processing, Pattern Recognition, and Intelligent Systems. Singapore: World Scientific Publishing Co; 2016.
63. Doshi A, Patrick K, Sallis JF, Calfas K. Evaluation of physical activity web sites for use of behavior change theories. Ann Behav Med 2003;25(2):105-111. [doi: 10.1207/S15324796ABM2502_06] [Medline: 12704012]
64. Bandura A. Social Foundations Of Thought And Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1986.
65. Fogg BJ. A Behavior Model for Persuasive Design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive'09; April 26-29, 2009; Claremont, California, USA. [doi: 10.1145/1541948.1541999]
66. 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]

67. Let's Move!. Healthy Communities URL: https://letsmove.obamawhitehouse.archives.gov/healthy-communities [WebCite Cache ID 78TbblnVJ]

## Abbreviations

**CNE:** culinary nutrition education
**FBM:** Fogg Behavior Model
**ICT:** information and communication technology
**SCT:** Social Cognitive Theory
**UCD:** user-centered design
**UTAUT:** Unified Theory of Acceptance and Use of Technology

*Edited by G Eysenbach; submitted 18.04.19; peer-reviewed by G De Leo, S Sittig; comments to author 14.05.19; revised version received 07.06.19; accepted 09.06.19; published 29.07.19.*

*Please cite as:*
*LeRouge C, Durneva P, Sangameswaran S, Gloster AM*
*Design Guidelines for a Technology-Enabled Nutrition Education Program to Support Overweight and Obese Adolescents: Qualitative User-Centered Design Study*
*J Med Internet Res 2019;21(7):e14430*
*URL: http://www.jmir.org/2019/7/e14430/*
*doi:10.2196/14430*
*PMID:31359871*

Original Paper

# Brief Motivational Interviewing Delivered by Clinician or Computer to Reduce Sexual Risk Behaviors in Adolescents: Acceptability Study

Taraneh Shafii[1], MPH, MD; Samantha K Benson[2], MPH; Diane M Morrison[3], PhD

[1]Division of Adolescent Medicine, Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, United States
[2]Harborview Medical Center, University of Washington School of Medicine, Seattle, WA, United States
[3]School of Social Work, University of Washington, Seattle, WA, United States

**Corresponding Author:**
Taraneh Shafii, MPH, MD
Division of Adolescent Medicine
Department of Pediatrics
University of Washington School of Medicine
CSB-200 PO Box 5371
Seattle, WA, 98145
United States
Phone: 1 206 987 2028
Fax: 1 206 987 3959
Email: taraneh.shafii@seattlechildrens.org

## Abstract

**Background:** Clinicians are expected to screen their adolescent patients for an increasing number of health behaviors and intervene when they uncover risky behaviors, yet, the clinic time allotted to screen, intervene, and provide resources is insufficient. Brief motivational interviewing (MI) offers succinct behavior change counseling; however, for implementation, clinicians need training, skill, and time. Computerized screening and counseling adjuvants may help clinicians increase their scope of behavioral screening, especially with sensitive topics such as sexual health, and provide risk-reduction interventions without consuming provider time during visits.

**Objective:** The objectives of this study were to (1) understand the extent to which health care providers use brief MI for sexual health discussions with adolescent patients and (2) assess the acceptability of incorporating a brief MI-based intervention to reduce sexual risk behaviors into their clinical practice delivered by either themselves or a computer.

**Methods:** At a national medical conference, surveys were administered to clinicians who provide sexual health care to adolescents. They were asked about their current use of MI for sexual risk behavior discussions and their willingness to implement computerized sexual health screening and computerized sexual risk behavior interventions into their clinical practice.

**Results:** The large majority (87.6%, 170/194) of clinicians already used MI with their patients with less than half (72/148, 48.6%) reporting they had been formally trained in MI. Despite all (195/195, 100.0%) clinicians feeling very or completely comfortable discussing sexual risk behaviors with their patients, the large majority (160/195, 82.1%) reported it would be useful, very useful, or extremely useful for a computerized program to do it all: screen their patients, generate risk profiles, and provide the risk-reduction counseling rather than doing it themselves.

**Conclusions:** In this study, most clinicians used some form of brief MI or client-centered counseling when discussing sexual risk behaviors with adolescents and are very comfortable doing so. However, the large majority would prefer to implement computerized sexual health screening, risk assessment, and sexual risk behavior interventions into their clinical care of adolescents.

***(J Med Internet Res 2019;21(7):e13220)*** doi:10.2196/13220

**KEYWORDS**

sexual health; risk behaviors; adolescent; healthcare providers; computer-assisted diagnosis; teen health; preventive care

## Introduction

### Background

Clinician sexual health discussions with adolescents remain suboptimal in real-world clinical practice [1-4]. Health care providers continue to search for optimal ways to communicate with adolescents about sexual health and risk behaviors. Some behavioral interventions have been shown to increase the knowledge of risk-reduction strategies (eg, condom and birth control use and negotiating safe sex with partners) and decrease self-reported unprotected sex; however, these interventions were tested in nonclinical settings and with specific populations of adolescents. Such interventions have yet to be tested or implemented in real-world outpatient settings and delivered by clinicians [5-9].

### Barriers to Screening

Even experienced clinicians in busy practices may not have the time to engage adolescents in discussions, which are needed to build rapport and uncover risk behaviors. Adolescents may have concerns about talking face-to-face with clinicians about sex or may not be granted enough time for confidential conversations during their visit [10-12]. Brief motivational interviewing (MI) has gained popularity as a means to engage adolescents in behavior change [13-20]; however, there are barriers to clinicians in adopting MI. It takes time to be trained and become proficient in MI, and effectively using MI requires already precious clinic visit time [21-22].

### Computer-Assisted Screening

Computerized screening with brief MI may serve to alleviate the time burden for health care providers and any discomfort in discussing sensitive health topics for both the clinician and patient. Computer screening improves adolescents' perceptions of medical visits [23-25]. The literature provides evidence that adolescents may prefer computerized sexual health screening to face-to-face interviews. A study of adolescents seeking care in a pediatric emergency department tested computerized sexual health screening and found that it was acceptable to adolescents, preferable to in-person interviews, and feasible for providers to implement in the emergency department [26]. A personal digital assistant screening tool that screened for several risk behaviors, including unprotected sex, was tested in primary care clinics before adolescent well visits and resulted in higher patient ratings for visit satisfaction, perceived confidentiality, and feeling listened to carefully [23].

### Computerized Interventions

Incorporating sexual behavior risk-reduction interventions into the computerized screening session takes these interventions one step further. Such interventions may be interactive and provide personalized feedback to the adolescent. Only a few computerized sexual health interventions for adolescents have been tested in real-world clinic settings, and these did not assess clinician acceptability of integrating the interventions into clinical practice [27-30]. Existing provider acceptability studies of computerized health screening and interventions are of adult patient populations, have small sample sizes, and may not include sexual health as a risk behavior [31-33]. We are not aware of any large studies assessing clinician willingness to be trained in brief MI for promoting adolescent sexual health. We were likewise unable to identify any studies assessing provider acceptability of incorporating computerized sexual health screening and interventions into visits with their adolescent patients. The objectives of this study were (1) to understand the extent to which health care providers use MI for sexual health with adolescent patients and (2) to assess the acceptability of incorporating a brief MI-based intervention to reduce sexual risk behaviors into their clinical practice delivered by either themselves or a computer.

## Methods

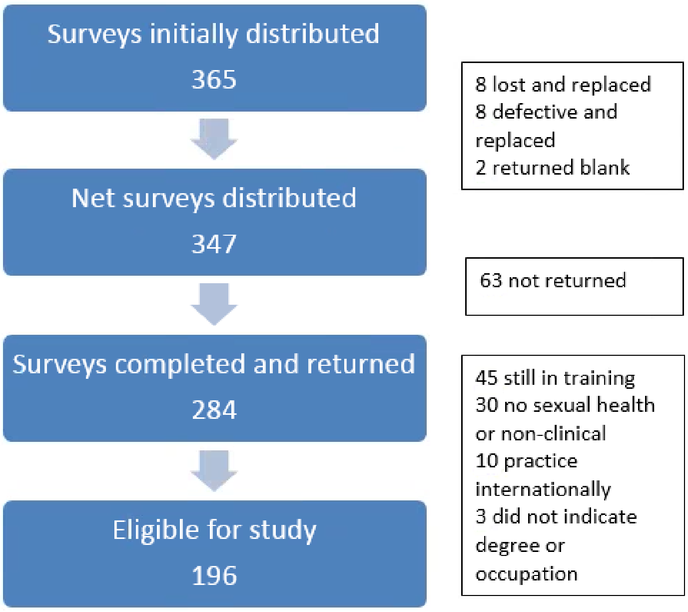
### Recruitment

In March 2009, we administered a 28-item survey to clinicians at a national medical conference, with attendees representing a wide geographic range in the United States. For the purposes of this study, clinicians were asked about sexually transmitted infection (STI) testing and positive STI diagnoses in the past 3 months to characterize their patient population and practice experience. The inclusion criteria were clinicians practicing in the United States who provided sexual and reproductive health care to adolescents. The exclusion criteria were not providing such care to adolescents or being in training. A total of 365 surveys were initially distributed and 18 were omitted from the final denominator (n=347) for the following reasons: the attendee returned the survey blank (n=8); the survey was lost by the participant and a replacement survey was provided (n=8); or the survey was defective because of printing error and was replaced with a corrected survey (n=2). Of the 347 surveys distributed, 81.8% (284/347) were completed. The University of Washington Human Subjects Division approved this study.

### Sample for Analysis

Of the 284 completed surveys, an additional 88 were ineligible because the attendee was in training (n=45), practicing outside the United States (n=10), not providing sexual and reproductive healthcare or not in practice or nonclinical (n=30), or did not indicate their degree or level of training (n=3). A total of 196 clinicians qualified for the study as they provided such care to adolescents, including the diagnosis and management of STIs and unintended pregnancy; identified themselves as medical doctor (MD)/doctor of osteopathy (DO), physician assistant (PA), or nurse practitioner (NP) and were not currently in training (Figure 1). STATA 11.0 by StataCorp LLC, was used for data analysis. *T* tests and chi-square tests were used to evaluate associations.

**Figure 1.** Flow diagram of sample for analysis.




## Results

### Demographic and Clinical Practice Characteristics

Physicians (MDs and DOs) comprised the majority of the sample (181/196, 92.3%), with (13/196, 6.6%) NPs and (2/196, 1.0%) PAs. In addition, 65.3% (128/196) of them were female. The most common practice type was academic (121/196, 61.7%) with a comparable proportion in private practice (19/196, 9.6%) and public health/community clinics (18/196, 9.2%). Over half (118/195, 60.1%) of the clinicians provided care only to adolescents (defined as ages 11 to 21 years). The number of years providers reported being in clinical practice ranged from less than 1 year to greater than 30 years and was evenly distributed across decades of practice. Through the use of Likert scales, we elicited information about clinical practice and patient population characteristics. The large majority (175/196, 89.3%) of clinicians provided care to 10 or more patients in a typical week. In the past 3 months, the majority (160/195, 82.1%) of clinicians tested 10 or more of their patients for STIs (*Neisseria gonorrhoeae*, *Chlamydia trachomatis*, trichomonas, genital warts, syphilis, and HIV) with the vast majority (178/196, 90.8%) of clinicians reporting 1 or more adolescents testing positive for an STI in the same time period (Table 1).

### Current Use of Motivational Interviewing in Clinical Practice

*All* clinicians reported that they felt at least very comfortable and a large majority felt *completely* comfortable (195/195, 100.0%; missing date, n=1) discussing sexual risk behaviors (eg, inconsistent or lack of condom use/hormonal birth control; multiple partners/concurrency; HIV/STI; unintended pregnancy) with their patients. Although all except 3 clinicians reported feeling that they were at least somewhat effective in changing sexual risk behaviors of their adolescent patients, only 22.1% (43/195) felt very or *completely* effective. Many clinicians reported they saw themselves as more effective at changing patient behaviors than other clinicians.

The vast majority of clinicians were familiar with MI defined in the survey as "… a directive, client-centered counseling style for eliciting behavior change by helping to explore and resolve ambivalence" [34]. The large majority of clinicians (170/194, 87.6%) already used MI with their patients. These 170 clinicians were asked if they were formally trained in MI and of 148 (missing data=22) respondents, less than half, 48.6% (72/148), reported formal training. Clinicians reported having used MI for many health topics, including obesity (93%), smoking (90%), alcohol (82%), substance abuse (87%), and sexual health (96%). Only half (52%) of the clinicians said they used it to discuss injury prevention (bike helmets/seat belts). Most clinicians (140/170, 82.4%) employ MI greater than half the time when discussing sexual health with their patients and feel they are more effective in communicating with their patient when they use MI compared with when they do not (Table 2).

### Motivational Interviewing Acceptability and Feasibility in Practice

The vast majority of clinicians would be willing to use MI or another type of client-centered counseling technique in their practice, if effective at reducing sexual risk behaviors in adolescents. Although 93% of them were willing to attend training for such an intervention, they preferred the length of training be limited to less than 1 day. Clinicians are willing to spend a maximum of 10 min per patient to deliver the intervention. Approximately half the clinicians (103/195, 52.8%) were willing to have follow-up contact with their patients as part of the sexual health intervention. However, 21.0% (41/195) of them would only do so if reimbursed. Clinicians were willing to do at least 1 monthly follow-up with their patients lasting less than 10 min per encounter. Preferred modes of follow-up in order of preference were telephone, email, text message, and social media (Table 3).

**Table 1.** Demographic characteristics of clinicians (N=196).

| Demographic characteristic | Statistics, n (%) |
|---|---|
| **Gender** | |
| Female | 128 (65.3) |
| Male | 68 (34.7) |
| **Clinician type** | |
| Medical doctor | 181 (92.4) |
| Nurse practitioner/physician assistant | 15 (7.6) |
| **Years in clinical practice (N=193), n (%)** | |
| ≤10 | 67 (34.7) |
| 11-20 | 64 (33.2) |
| >20 | 62 (32.1) |
| **Patient type (N=195), n (%)** | |
| Adolescents only (aged 11-21 years) | 118 (60.5) |
| Children and adolescents (aged 0-21 years) | 30 (15.4) |
| All ages (0 through adulthood) | 19 (9.7) |
| Adolescents and adults (aged 11 years through adult) | 28 (14.4) |
| **Practice type (N=195), n (%)** | |
| Academic | 121 (62.0) |
| Private | 20 (10.3) |
| Community/public health | 23 (11.8) |
| Other | 31 (15.9) |
| **Adolescent patients per week** | |
| <10 | 21 (10.7) |
| 10-29 | 87 (44.4) |
| ≥30 | 88 (44.9) |
| **STI[a] tests on patients per 3 months (N=195), n (%)** | |
| <10 | 35 (17.9) |
| 10-29 | 59 (30.2) |
| ≥30 | 101 (51.9) |
| **Positive STI[a] tests per 3 months** | |
| <10 | 123 (62.7) |
| 10-29 | 55 (28.1) |
| ≥30 | 18 (9.2) |

[a]STI: sexually transmitted infection, including chlamydia, gonorrhea, trichomonas, herpes, genital warts, syphilis, and HIV.

**Table 2.** Clinician perspective of sexual risk behaviors and motivational interviewing (MI; N=196).

| Clinician perspectives | Statistics, n (%) |
|---|---|
| **Comfort talking about sexual risk behaviors[a] (N=195), n (%)** | |
| Not comfortable | 0 (0) |
| Somewhat comfortable | 0 (0) |
| Comfortable | 0 (0) |
| Very comfortable | 27 (13.8) |
| Completely comfortable | 168 (86.2) |
| **Clinician effectiveness in changing sexual risk behaviors (N=194), n (%)** | |
| Not effective | 8 (4.1) |
| Somewhat effective | 89 (45.9) |
| Effective | 68 (35.0) |
| Very effective | 29 (15.0) |
| Completely effective | 0 (0) |
| **Personal effectiveness in changing behavior (N=195), n (%)** | |
| Not effective | 3 (1.5) |
| Somewhat effective | 76 (39.0) |
| Effective | 73 (37.5) |
| Very effective | 42 (21.5) |
| Completely effective | 1 (0.5) |
| **Use of MI with patient** | |
| Yes | 170 (87.6) |
| No | 24 (12.4) |
| **Formally trained in MI (N=170)[a,b] (n=148), n (%)** | |
| Yes | 72 (48.6) |
| No | 76 (51.4) |
| **Types of behavioral issues addressed** | |
| Sexual risk behavior | 163 (96.4) |
| Obesity | 155 (92.8) |
| Smoking cigarettes | 147 (89.6) |
| Drinking alcohol | 133 (82.1) |
| Substance abuse | 140 (87.0) |
| Injury prevention (bike helmets/seat belts) | 83 (52.2) |
| **Frequency of use of MI with patients for sexual risk behavior** | |
| Never | 2 (1.2) |
| 25% of time | 28 (16.5) |
| 50% of time | 41 (24.1) |
| 75% of time | 53 (31.2) |
| Almost always | 46 (27.0) |
| **Provider effectiveness in changing behavior when using MI versus when not (N=163), n (%)** | |
| Much less effective | 1 (0.6) |
| Somewhat less effective | 4 (2.4) |
| No difference | 22 (13.5) |

| Clinician perspectives | Statistics, n (%) |
|---|---|
| More effective | 120 (73.6) |
| Much more effective | 16 (9.8) |

[a]No difference by number practice years or frequency sexually transmitted infection testing.

[b]Remaining survey questions only asked if ever used motivational interviewing, n=170.

**Table 3.** Clinician perspective of self-delivered motivational interviewing (MI).

| Clinician perspective | Statistics |
|---|---|
| **Willing to attend MI training (N=186), n (%)** | |
| Yes | 174 (93.6) |
| No | 12 (6.4) |
| **Maximum length training (N=177), n (%)** | |
| ≤2 hours | 22 (12.4) |
| Half day | 58 (32.8) |
| 1 day | 54 (30.5) |
| ≥2 days | 43 (24.3) |
| **Maximum length of MI session with patient (N=192), n (%)** | |
| ≤5 min | 16 (8.3) |
| 5 min | 62 (32.3) |
| 10 min | 67 (34.9) |
| 15 min | 30 (15.6) |
| ≥20 min | 17 (8.9) |
| **Feasible for clinician follow-up with patient (N=195), n (%)** | |
| Yes | 103 (52.8) |
| No | 51 (26.2) |
| Only if reimbursed | 41 (21.0) |
| **Maximum length of MI follow-up session (N=152), n (%)** | |
| ≤5 min | 99 (65.1) |
| 10 min | 31 (20.4) |
| 15 min | 12 (7.9) |
| ≥20 min | 10 (6.6) |
| **Maximum number of monthly follow-up contacts (N=158), n (%)** | |
| 1 | 30 (19.0) |
| 2 | 39 (24.7) |
| 3 | 23 (14.6) |
| 4-5 | 29 (18.3) |
| 6 | 37 (23.4) |
| **Follow-up method willing to use, n/N (%)** | |
| Phone call | 122/149 (81.9) |
| Text message | 90/144 (62.5) |
| Email | 124/149 (83.2) |
| Social media | 30/139 (21.6) |

### Computer-Delivered Risk-Reduction Acceptability

The large majority of clinicians (165/192, 85.9%) found it more feasible for a computer to provide the sexual risk behavior screening intervention to their patients rather than themselves and would use a computer-generated sexual risk profile printout to facilitate discussion with their patients. The large majority of clinicians (160/195, 82.1%) also thought it would be useful, very useful, or *extremely* useful for the computer do it all: screen their patient, generate their sexual risk profile, and provide the risk-reduction counseling itself, requiring the provider to review only the findings with their patients afterward.

### Preference for Computerized Risk Screening and Risk-Reduction Counseling

No associations were found when comparing the number of years in clinical practice and comfort discussing sexual risk behaviors with adolescents; being trained in MI; or preferring computerized sexual risk screening and risk-reduction counseling. There was also no association between preference for computerized risk screening and counseling by clinician gender, type of practice, number of patients seen per week, and number of patients tested or testing positive for an STI in the past 3 months (Table 4).

**Table 4.** Clinician perspective of motivational interviewing (MI) and computer-delivered risk reduction for sexual health (N=196).

| Clinician perspective | Statistics, n (%) |
|---|---|
| **If MI sexual behavior risk reduction effectively delivered via clinician would it be feasible for you to do yourself?[a,b]** | |
| Yes | 183 (95.8) |
| No | 8 (4.2) |
| **If sexual behavior risk reduction effectively delivered via computer would that be more feasible for you than doing it yourself?[b,c]** | |
| Yes | 165 (85.9) |
| No | 27 (14.1) |
| **Likeliness to use computer printout of sexual risk behavior profile to facilitate risk-reduction counseling** | |
| Not likely | 8 (4.1) |
| Somewhat likely | 34 (17.3) |
| Likely | 59 (30.1) |
| Very likely | 68 (34.7) |
| Extremely likely | 27 (13.8) |
| **How useful would it be for you if computer generated a printout of sexual risk behavior profile AND provided risk-reduction counseling requiring you to do nothing further OR to simply review the findings with your adolescent patients?[b,d]** | |
| Not useful | 7 (3.6) |
| Somewhat useful | 28 (14.4) |
| Useful | 53 (27.2) |
| Very useful | 59 (30.3) |
| Extremely useful | 48 (24.6) |

[a]n=191.

[b]No difference by number of years in clinical practice or frequency of sexually transmitted infection testing.

[c]n=192.

[d]n=195.

## Discussion

### Principal Findings

In a survey of clinicians who provide sexual health care to adolescents from varied geographic regions around the United States and a wide range of clinical experience, the vast majority reported being very comfortable discussing sexual health with their adolescent patients. The majority of clinicians reported using MI for sexual health counseling with their patients, although less than half of these reported formally training in MI. Surprisingly, this sample of clinicians espousing such comfort with adolescent sexual health discussions reported that it would be preferable for a computer to do it all: screen for sexual risk behaviors and provide their patients with risk-reduction counseling.

### Comparison With Previous Work

There has been increasing focus on sexual health screening and MI in medical school curricula over the past 2 decades [35-40]. Other studies have found younger clinicians to be more comfortable discussing sexual health and using MI for behavior change as compared with older providers [41,3,1]. However, in this survey, providers with more than 30 years of clinical experience were just as likely as those with less than 10 years of clinical experience to report comfort in talking about sexual health and using MI with patients. The similarity in comfort across respondents with different practice longevity could be because of most clinicians in the study primarily taking care of adolescents and so were comfortable with the population. Also most clinicians were in academic practice, and may be early adopters of evolving clinical practice approaches over the years.

Most providers felt they were at least somewhat effective at influencing the sexual behavior of their patients. This sentiment echoes a qualitative study, with physicians reporting they had influence in the choice of contraception with their female patients [42]. In our study, most respondents considered themselves more effective when using MI than when not, and the majority of them considered themselves more effective at encouraging patients' behavior change compared with other clinicians. Although there were no studies found in the literature that addressed providers' perceptions of their effectiveness with MI, there are existing studies that demonstrate clinician use of MI for behavior change to be efficacious in changing health risk behaviors [13-18].

Although respondents were very comfortable discussing sex with adolescents and even felt they were effective at eliciting behavior change, they considered it more feasible for a computer to administer the screening *and* counseling rather than doing so themselves. To our knowledge, this finding is novel in the literature. For this population of providers, preferring a computerized approach to sexual health risk-reduction counseling may reflect time limitations for patient visits rather than reticence to discuss sexual health. Most providers were willing to be trained in an MI sexual risk behavior intervention that includes at least 1 follow-up session; however, 20% of providers indicated they would only do a follow-up session with patients if they were reimbursed, which may also reflect increasing pressures on clinicians for productivity.

### Limitations

A limitation of this study is that we did not define MI in detail or what is required for training and proficiency in true MI. In the survey, we defined MI as "… a directive, client-centered counseling style for eliciting behavior change by helping clients explore and resolve ambivalence" [34]. It is possible that participants have different definitions for and experience in the use of MI, which may have biased the responses to questions about the use of and training in MI. In addition, the proposed computerized screening and intervention was theoretical, so clinicians were not providing feedback on a tangible product for which they may have different opinions. As most clinicians practiced in academic settings, the findings may not be generalizable to clinicians in other practice types. The decision was made to focus on clinicians practicing in the United States to account for the large variation worldwide in attitudes toward adolescent sexual and reproductive health and clinical practice. The authors acknowledge that this was a missed opportunity to learn about international clinician practice.

MI has gained increasing popularity over the past decade since this study. *Brief* MI is used for many different health behaviors, and we anticipate an even higher acceptability by medical providers. However, the issue of lack of provider time with patients has also escalated over the past decade. These data are relevant as providers have not yet found an answer and continue to strategize on how they can provide comprehensive health care in the limited minutes they have for adolescent patient visits. Such an intervention as presented in this study is a possible solution.

### Conclusions

Clinicians are increasingly pressed for time when providing care to patients and researchers and practitioners have not yet found the most effective way to consistently discuss sexual health with adolescents or promote healthy sexual behaviors. Computerized interventions, which incorporate both behavioral screening and risk-reduction counseling, may provide solutions to both issues. The development of computerized health interventions is a rapidly growing field and further research is needed to create and test such interventions in real-world clinical practice.


## Acknowledgments

The authors would like to acknowledge the Funding Agency: Eunice Kennedy Shriver National Institute of Child Health and Human Development NICHD Grant No: 5K23HD052621, eGC1 Number: A66661; Project Title: Brief Clinician Intervention for High-Risk Behaviors in Adolescents, Grant Principal Investigator: TS; Project Period Dates: 7/1/07-5/31/13. All of those who have contributed significantly to this study are listed. This study was presented as a poster: TS, DMM, MRG, and KKH. "Acceptability to health care providers of using motivational interviewing delivered via clinician or computer in clinical practice to reduce sexual risk behaviors in adolescents," at the 2013 Annual Meeting of the Society for Adolescent Health and Medicine, Atlanta, Georgia March 2013.


## Conflicts of Interest

None declared.

## References


1. Alexander SC, Fortenberry JD, Pollak KI, Bravender T, Davis JK, Ostbye T, et al. Sexuality talk during adolescent health maintenance visits. JAMA Pediatr 2014 Feb;168(2):163-169 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.4338] [Medline: 24378686]

2. Marcell AV, Bell DL, Lindberg LD, Takruri A. Prevalence of sexually transmitted infection/human immunodeficiency virus counseling services received by teen males, 1995-2002. J Adolesc Health 2010 Jun;46(6):553-559 [FREE Full text] [doi: 10.1016/j.jadohealth.2009.12.002] [Medline: 20472212]
3. Kelts EA, Allan MJ, Klein JD. Where are we on teen sex?: Delivery of reproductive health services to adolescents by family physicians. Fam Med 2001 May;33(5):376-381. [Medline: 11355649]
4. Blum RW, Beuhring T, Wunderlich M, Resnick MD. Don't ask, they won't tell: the quality of adolescent health screening in five practice settings. Am J Public Health 1996 Dec;86(12):1767-1772. [Medline: 9003135]
5. Jemmott JB, Jemmott LS, Fong GT, Morales KH. Effectiveness of an HIV/STD risk-reduction intervention for adolescents when implemented by community-based organizations: a cluster-randomized controlled trial. Am J Public Health 2010 Apr;100(4):720-726. [doi: 10.2105/AJPH.2008.140657] [Medline: 20167903]
6. Jemmott JB, Jemmott LS, O'Leary A, Ngwane Z, Icard LD, Bellamy SL, et al. School-based randomized controlled trial of an HIV/STD risk-reduction intervention for South African adolescents. Arch Pediatr Adolesc Med 2010 Oct;164(10):923-929 [FREE Full text] [doi: 10.1001/archpediatrics.2010.176] [Medline: 20921349]
7. DiClemente RJ, Wingood GM, Harrington KF, Lang DL, Davies SL, Hook EW, et al. Efficacy of an HIV prevention intervention for African American adolescent girls: a randomized controlled trial. J Am Med Assoc 2004 Jul 14;292(2):171-179. [doi: 10.1001/jama.292.2.171] [Medline: 15249566]
8. Heeren GA, Jemmott JB, Ngwane Z, Mandeya A, Tyler JC. A randomized controlled pilot study of an HIV risk-reduction intervention for sub-Saharan African university students. AIDS Behav 2013 Mar;17(3):1105-1115 [FREE Full text] [doi: 10.1007/s10461-011-0129-2] [Medline: 22246515]
9. Morrison-Beedy D, Jones SH, Xia Y, Tu X, Crean HF, Carey MP. Reducing sexual risk behavior in adolescent girls: results from a randomized controlled trial. J Adolesc Health 2013 Mar;52(3):314-321 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.07.005] [Medline: 23299011]
10. McKee M, Fletcher J, Schechter CB. Predictors of timely initiation of gynecologic care among urban adolescent girls. J Adolesc Health 2006 Aug;39(2):183-191. [doi: 10.1016/j.jadohealth.2005.11.022] [Medline: 16857529]
11. Hoopes AJ, Benson SK, Howard HB, Morrison DM, Ko LK, Shafii T. Adolescent perspectives on patient-provider sexual health communication: a qualitative study. J Prim Care Community Health 2017 Oct;8(4):332-337 [FREE Full text] [doi: 10.1177/2150131917730210] [Medline: 28929860]
12. Fairbrother G, Scheinmann R, Osthimer B, Dutton MJ, Newell KA, Fuld J, et al. Factors that influence adolescent reports of counseling by physicians on risky behavior. J Adolesc Health 2005 Dec;37(6):467-476. [doi: 10.1016/j.jadohealth.2004.11.001] [Medline: 16310124]
13. Gayes LA, Steele RG. A meta-analysis of motivational interviewing interventions for pediatric health behavior change. J Consult Clin Psychol 2014 Jun;82(3):521-535. [doi: 10.1037/a0035917] [Medline: 24547922]
14. Brown RA, Abrantes AM, Minami H, Prince MA, Bloom EL, Apodaca TR, Hunt JI. Motivational Interviewing to Reduce Substance Use in Adolescents with Psychiatric Comorbidity. J Subst Abuse Treat 2015;59:20-29. [Medline: 26362000]
15. Stanger C, Ryan SR, Delhey LM, Thrailkill K, Li Z, Budney AJ. A multicomponent motivational intervention to improve adherence among adolescents with poorly controlled type 1 diabetes: a pilot study. J Pediatr Psychol 2013 Jul;38(6):629-637 [FREE Full text] [doi: 10.1093/jpepsy/jst032] [Medline: 23699750]
16. Gourlan M, Sarrazin P, Trouilloud D. Motivational interviewing as a way to promote physical activity in obese adolescents: a randomised-controlled trial using self-determination theory as an explanatory framework. Psychol Health 2013 Nov;28(11):1265-1286. [doi: 10.1080/08870446.2013.800518] [Medline: 23756082]
17. Walton M, Chermack ST, Shope JT, Bingham CR, Zimmerman MA, Blow FC, et al. Effects of a brief intervention for reducing violence and alcohol misuse among adolescents: a randomized controlled trial. J Am Med Assoc 2010 Aug 4;304(5):527-535 [FREE Full text] [doi: 10.1001/jama.2010.1066] [Medline: 20682932]
18. Cunningham RM, Chermack ST, Zimmerman MA, Shope JT, Bingham CR, Blow FC, et al. Brief motivational interviewing intervention for peer violence and alcohol use in teens: one-year follow-up. Pediatrics 2012 Jun;129(6):1083-1090 [FREE Full text] [doi: 10.1542/peds.2011-3419] [Medline: 22614776]
19. Gold MA, Tzilos GK, Stein LA, Anderson BJ, Stein MD, Ryan CM, et al. A randomized controlled trial to compare computer-assisted motivational intervention with didactic educational counseling to reduce unprotected sex in female adolescents. J Pediatr Adolesc Gynecol 2016 Feb;29(1):26-32 [FREE Full text] [doi: 10.1016/j.jpag.2015.06.001] [Medline: 26514957]
20. Barnet B, Liu J, DeVoe M, Duggan AK, Gold MA, Pecukonis E. Motivational intervention to reduce rapid subsequent births to adolescent mothers: a community-based randomized trial. Ann Fam Med 2009;7(5):436-445 [FREE Full text] [doi: 10.1370/afm.1014] [Medline: 19752472]
21. McKee MD, Rubin SE, Campos G, O'Sullivan LF. Challenges of providing confidential care to adolescents in urban primary care: clinician perspectives. Ann Fam Med 2011;9(1):37-43 [FREE Full text] [doi: 10.1370/afm.1186] [Medline: 21242559]
22. Emmons KM, Rollnick S. Motivational interviewing in health care settings. Opportunities and limitations. Am J Prev Med 2001 Jan;20(1):68-74. [Medline: 11137778]

23. Olson AL, Gaffney CA, Hedberg VA, Gladstone GR. Use of inexpensive technology to enhance adolescent health screening and counseling. Arch Pediatr Adolesc Med 2009 Feb;163(2):172-177. [doi: 10.1001/archpediatrics.2008.533] [Medline: 19188650]
24. Brown JD, Wissow LS. Discussion of sensitive health topics with youth during primary care visits: relationship to youth perceptions of care. J Adolesc Health 2009 Jan;44(1):48-54 [FREE Full text] [doi: 10.1016/j.jadohealth.2008.06.018] [Medline: 19101458]
25. Zieve GG, Richardson LP, Katzman K, Spielvogle H, Whitehouse S, McCarty CA. Adolescents' perspectives on personalized e-feedback in the context of health risk behavior screening for primary care: qualitative study. J Med Internet Res 2017 Dec 20;19(7):e261 [FREE Full text] [doi: 10.2196/jmir.7474] [Medline: 28729236]
26. Goyal MK, Shea JA, Hayes KL, Badolato G, Chamberlain JM, Zaoutis T, et al. Development of a sexual health screening tool for adolescent emergency department patients. Acad Emerg Med 2016 Dec;23(7):809-815 [FREE Full text] [doi: 10.1111/acem.12994] [Medline: 27126128]
27. Peipert J, Redding CA, Blume J, Allsworth JE, Iannuccillo K, Lozowski F, et al. Design of a stage-matched intervention trial to increase dual method contraceptive use (Project PROTECT). Contemp Clin Trials 2007 Sep;28(5):626-637. [doi: 10.1016/j.cct.2007.01.012] [Medline: 17374567]
28. Peipert JF, Redding CA, Blume JD, Allsworth JE, Matteson KA, Lozowski F, et al. Tailored intervention to increase dual-contraceptive method use: a randomized trial to reduce unintended pregnancies and sexually transmitted infections. Am J Obstet Gynecol 2008 Jun;198(6):630.e1-630.e8 [FREE Full text] [doi: 10.1016/j.ajog.2008.01.038] [Medline: 18395692]
29. Downs JS, Murray PJ, de Bruin WB, Penrose J, Palmgren C, Fischhoff B. Interactive video behavioral intervention to reduce adolescent females' STD risk: a randomized controlled trial. Soc Sci Med 2004 Oct;59(8):1561-1572. [doi: 10.1016/j.socscimed.2004.01.032] [Medline: 15279915]
30. DiClemente RJ, Wingood GM, Rose ES, Sales JM, Lang DL, Caliendo AM, et al. Efficacy of sexually transmitted disease/human immunodeficiency virus sexual risk-reduction intervention for african american adolescent females seeking sexual health services: a randomized controlled trial. Arch Pediatr Adolesc Med 2009 Dec;163(12):1112-1121. [doi: 10.1001/archpediatrics.2009.205] [Medline: 19996048]
31. Mackenzie SL, Kurth AE, Spielberg F, Severynen A, Malotte CK, St Lawrence J, et al. Patient and staff perspectives on the use of a computer counseling tool for HIV and sexually transmitted infection risk reduction. J Adolesc Health 2007 Jun;40(6):572.e9-572.16. [doi: 10.1016/j.jadohealth.2007.01.013] [Medline: 17531766]
32. Paul CL, Carey M, Yoong SL, D'Este C, Makeham M, Henskens F. Access to chronic disease care in general practice: the acceptability of implementing systematic waiting-room screening using computer-based patient-reported risk status. Br J Gen Pract 2013 Sep;63(614):e620-e626 [FREE Full text] [doi: 10.3399/bjgp13X671605] [Medline: 23998842]
33. Ahmad F, Skinner HA, Stewart DE, Levinson W. Perspectives of family physicians on computer-assisted health-risk assessments. J Med Internet Res 2010 May 7;12(2):e12 [FREE Full text] [doi: 10.2196/jmir.1260] [Medline: 20457555]
34. Rollnick S, Miller WR. What is motivational interviewing? Behav Cogn Psychother 2009 Jun 16;23(4):325-334. [doi: 10.1017/S135246580001643X]
35. Coleman E. Sexual health education in medical school: a comprehensive curriculum. Virtual Mentor 2014 Nov 1;16(11):903-908. [doi: 10.1001/virtualmentor.2014.16.11.medu1-1411] [Medline: 25397650]
36. Bayer CR, Eckstrand KL, Knudson G, Koehler J, Leibowitz S, Tsai P, et al. Sexual health competencies for undergraduate medical education in North America. J Sex Med 2017 Dec;14(4):535-540. [doi: 10.1016/j.jsxm.2017.01.017] [Medline: 28202322]
37. Galletly C, Lechuga J, Layde JB, Pinkerton S. Sexual health curricula in U.S. medical schools: current educational objectives. Acad Psychiatry 2010;34(5):333-338. [doi: 10.1176/appi.ap.34.5.333] [Medline: 20833900]
38. Poirier MK, Clark MM, Cerhan JH, Pruthi S, Geda YE, Dale LC. Teaching motivational interviewing to first-year medical students to improve counseling skills in health behavior change. Mayo Clin Proc 2004 Mar;79(3):327-331. [doi: 10.4065/79.3.327] [Medline: 15008606]
39. Haeseler F, Fortin AH, Pfeiffer C, Walters C, Martino S. Assessment of a motivational interviewing curriculum for year 3 medical students using a standardized patient case. Patient Educ Couns 2011 Jul;84(1):27-30 [FREE Full text] [doi: 10.1016/j.pec.2010.10.029] [Medline: 21123019]
40. White LL, Gazewood JD, Mounsey AL. Teaching students behavior change skills: description and assessment of a new Motivational interviewing curriculum. Med Teach 2007 May;29(4):e67-e71. [doi: 10.1080/01421590601032443] [Medline: 17786734]
41. Henry-Reid LM, O'Connor KG, Klein JD, Cooper E, Flynn P, Futterman DC. Current pediatrician practices in identifying high-risk behaviors of adolescents. Pediatrics 2010;125(4):e741-e747 [FREE Full text]
42. Henderson JT, Raine T, Schalet A, Blum M, Harper CC. "I wouldn't be this firm if I didn't care": preventive clinical counseling for reproductive health. Patient Educ Couns 2011 Feb;82(2):254-259 [FREE Full text] [doi: 10.1016/j.pec.2010.05.015] [Medline: 20558024]

## Abbreviations

**DO:** doctor of osteopathy
**MD:** medical doctor
**MI:** motivational interviewing
**NP:** nurse practitioner
**PA:** physician assistant
**STI:** sexually transmitted infection

*Edited by G Eysenbach; submitted 21.12.18; peer-reviewed by J Torres, A Aventin; comments to author 14.02.19; revised version received 05.04.19; accepted 23.04.19; published 10.07.19.*

*Please cite as:*
*Shafii T, Benson SK, Morrison DM*
*Brief Motivational Interviewing Delivered by Clinician or Computer to Reduce Sexual Risk Behaviors in Adolescents: Acceptability Study*
*J Med Internet Res 2019;21(7):e13220*
*URL: http://www.jmir.org/2019/7/e13220/*
*doi:10.2196/13220*
*PMID:*

Original Paper

# The PrEP You Want: A Web-Based Survey of Online Cross-Border Shopping for HIV Prophylaxis Medications

Ben Walmsley[1], BA; Dan Gallant[2], BA, MA; Mark Naccarato[1], BPharm; Mark Hull[3], BSc, MSc, MD; Alex Smith[4], BSN; Darrell Hoi-San Tan[1], BSc, MD, PhD

[1]Division of Infectious Diseases, St Michael's Hospital, Toronto, ON, Canada
[2]Gay Men’s Sexual Health Alliance, Toronto, ON, Canada
[3]Division of Infectious Diseases, University of British Columbia, Vancouver, BC, Canada
[4]Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

**Corresponding Author:**
Ben Walmsley, BA
Division of Infectious Diseases
St Michael's Hospital
209 Victoria Street
Toronto, ON, M5B 1T8
Canada
Phone: 1 416 864 6060 ext 77105
Email: ben.j.walmsley@gmail.com

## Abstract

**Background:** In response to the high cost of HIV pre-exposure prophylaxis (PrEP) medications in Canada, community organizations have created internet-based guides detailing how to legally order generic medications online and travel to collect them in the United States. However, little is known about the patients following these guides.

**Objective:** Our primary objective was to measure the proportion of Ontario gay, bisexual, and other men who have sex with men (GBMSM) accessing these online guides who intended to use the border-crossing approach. Our secondary objectives were to explore their demographic characteristics, their completion of the steps in the border-crossing approach, and the barriers they perceived.

**Methods:** Between July 20, 2017, and May 18, 2018, we administered two online surveys of GBMSM accessing an online border-crossing guide posted by a gay men’s health organization in Ontario. Participants completed an open baseline survey posted on the border-crossing guide’s Web page and a follow-up survey 3 months later. The data were analyzed using descriptive statistics. We used multivariable logistic regression to identify characteristics associated with the intention to use the border-crossing approach.

**Results:** Most of the 141 participants were young (median age 23, interquartile range 22-25 years) and black (79.4%; 112/141) GBMSM who had completed a college or an undergraduate degree (62.4%; 88/141). In addition, 19.9% (28/141) of them reported a total family income less than Can $30,000 and another 53.9% (76/141) reported income between Can $30,000 and Can $60,000. 54.6% (76/141) paid for medications entirely out of pocket. Most participants indicated that they were likely to complete a border-crossing approach: 80.1% (113/141) at baseline and 79.1% (87/110) at follow-up. The characteristics associated with the intention to use the approach included being black (adjusted odds ratio [aOR] 5.73, 95% CI 2.06-16.61), paying for medications out of pocket (aOR 5.18, 95% CI 1.82-17.04), and having a provider who was thought to be willing to prescribe PrEP (aOR 4.42, 95% CI 1.63-12.41). Comparing baseline and follow-up for the 110 participants who completed both surveys, 65.4% (72/110) and 80.0% (88/110) had discussed PrEP with a health care provider, 18.1% (20/110) and 25.4% (28/110) had obtained a PrEP prescription, and 8.2% (9/110) and 5.5% (6/110) had ordered medications to that mailbox, whereas only 1.0% (1/110) and 0.0% (0/110) had crossed the border to collect them at baseline and follow-up, respectively. Reported barriers included perceived concerns about the approach’s legality (56.0%; 79/141), the security of personal health information (39.0%; 55/141), and the safety of online vendors (38.3%; 54/141).

**Conclusions:** Despite high interest in pursuing an online border-crossing approach to get PrEP medications, such an approach may not be a viable option for PrEP scale-up among interested GBMSM because of logistical challenges and perceptions of safety and legitimacy.

***(J Med Internet Res 2019;21(7):e12076)*** doi:10.2196/12076

**KEYWORDS**

pre-exposure prophylaxis; tenofovir disoproxil fumarate/emtricitabine; generic antiretroviral drugs; online medication shopping; men who have sex with men; HIV

## Introduction

### Background

HIV infection rates remain high among gay, bisexual, and other men who have sex with men (GBMSM) living in Canada, accounting for an estimated 47.9% of the nation's 2328 new adult cases in 2016 [1]. Pre-exposure prophylaxis (PrEP) is a promising strategy for preventing HIV infection [2]. Pharmacokinetic models suggest that consistent use of PrEP, which currently involves regular oral dosing of two coformulated antiretroviral medications, tenofovir disoproxil fumarate and emtricitabine (TDF/FTC), reduces the risk of HIV infection by almost 100% in GBMSM [3-5]. Furthermore, we have documented high and increasing willingness to use PrEP among GBMSM living in Canada. Although 33.3% of a sample of GBMSM attending a major Toronto sexual health clinic for HIV testing in 2011 said they would definitely be willing to use PrEP, this proportion increased to 52.5% in 2015 [6].

Despite this interest, the cost of PrEP medications has been a barrier to its uptake in Canada, where health services are publicly funded but medication costs are not universally covered. Brand name TDF/FTC (Truvada) costs approximately Can $876 per month in Ontario [7], and although generic TDF/FTC is less expensive at approximately Can $220 per month, this cost remains excessive for many. PrEP is publicly subsidized in some but not all Canadian jurisdictions at present, and in some cases, including Ontario, coverage may be only partial, requiring out-of-pocket copayments. Accordingly, the cost of PrEP is still prohibitive for many Canadians [8]. PrEP access in other high-income jurisdictions is variable, although notable examples of truly universal PrEP coverage have begun to emerge, including in France, Scotland, and New South Wales.

In response to this barrier in Canada, motivated patients have been pursuing alternate strategies to obtain less expensive generic PrEP medication, including crossing the border to the United States to pick up PrEP that they have ordered online [9]. Although having prescription drugs shipped directly into Canada for personal use is illegal [10], individuals crossing the border can legally import up to 90 days' worth of medication for personal use. Therefore, innovative activists and community organizations have created online resources detailing how a border-crossing approach can be used to legally acquire more affordable PrEP. The process involves four basic steps: (1) getting a prescription, (2) getting a mailbox in a US border location, (3) ordering medication online that ships to that location, and (4) picking up medication from that US location and crossing the border back to Canada every 3 months as needed. This approach is legal and, thus, its safety should be similar to crossing the border for any other reason. Reports from community members and clinicians alike indicate that this approach has been widely used in British Columbia [9,11], and a new campaign began promoting the strategy in Ontario in mid-2017. However, there are few data on the characteristics and motivations of the strategy's users in any setting and any perceived or actual barriers they may face.

### Objectives

To address these gaps, we conducted two Web-based surveys, one baseline and one follow-up, of GBMSM who were considering following the border-crossing approach as described by the major Ontario community organization championing the new campaign, the Gay Men's Sexual Health Alliance (GMSH), in an online resource titled *The PrEP You Want* [12]. Our primary objective was to measure the proportion of individuals accessing the resource who indicated an intention to use the border-crossing approach. Our secondary objectives were to explore their characteristics, their completion of the steps in the border-crossing process (Textbox 1), and the barriers they perceived.

**Textbox 1.** Steps in a cascade of online shopping and border crossing to obtain pre-exposure prophylaxis (PrEP).

1. Has decided to get PrEP
2. Has discussed PrEP with a health care provider
3. Has found a health care provider who will prescribe PrEP
4. Has decided to obtain PrEP with the border-crossing strategy
5. Has obtained a prescription from a health care provider
6. Has secured a mailbox in a US city
7. Has ordered medication online for shipment to that US mailbox
8. Has picked up PrEP medication from the US mailbox and crossed the border back to Canada

## Methods

### Study Design

Between July 20, 2017, and May 18, 2018, we administered two online surveys, one baseline and one follow-up, of GBMSM considering the border-crossing approach. The study was an open survey: to enter the study, anyone who self-reported that they met the eligibility criteria could access it by clicking on an online study link. We posted the link on the GMSH online information page, and the GMSH advertised this information page on several websites (eg, AIDS Committee of Toronto website, *PrEP—Canada: Rethinking HIV Prevention* Facebook page). We sent all the eligible participants a baseline questionnaire along with a unique participant study code by email, and after 3 months, we sent a follow-up questionnaire to capture updated information on their experience. For both the baseline and the follow-up surveys, we contacted participants who had not yet completed their questionnaires two more times with reminders.

### Participants

The participants were eligible if they self-identified as being gay, bisexual, or otherwise a man who has sex with men; were residents of Ontario; were first-time participants in the study; and were able to read and write English. In addition, they had to report having read the GMSH online information page and having determined whether they were likely to use the strategy or not. All participants were offered a Can $5 electronic gift card on completion of the baseline questionnaire, and a Can $10 electronic gift card on completion of the follow-up questionnaire.

### Survey Instrument

We designed both the 37-item baseline questionnaire and the 27-item follow-up questionnaire based on previous surveys of potential PrEP users [13] and the language of the GMSH online resource. The questionnaires covered the following domains: demographics; HIV risk behaviors; health care access; knowledge and experience of PrEP; and interest and experience with the border-crossing strategy. We developed the demographic section to match a standardized health equity questionnaire developed by major health care organizations in Toronto [14]. The HIV risk section included all items from the High Incidence Risk Index for men who have sex with men (HIRI-MSM), a HIV risk index developed and validated by the US Centers for Disease Control for its use in predicting seroconversion in GBMSM [15]. It should be noted that although the HIRI-MSM questions are based on 6-month look-back intervals, our questionnaires used 3-month intervals to avoid overlap between the baseline and follow-up responses; these 3-month responses were then doubled to estimate the final HIRI-MSM scores. We housed both surveys on the website *HostedinCanadaSurveys*.

### Analysis

Data were primarily analyzed with descriptive statistics, and the analysis was conducted separately for the baseline data and the follow-up data. For the primary objective, we calculated the proportion of individuals who answered *likely* or *very likely* to the question: “How likely is it that you will use this ‘The PrEP You Want: How to Order PrEP Online’ approach to buy generic PrEP drugs?”. For the secondary objectives, we calculated the proportion of individuals who reported completing each of the cascade of steps in Textbox 1 and the proportion reporting a variety of barriers to carrying out *The PrEP You Want* approach.

As an exploratory analysis, we constructed univariable and multivariable logistic regression models quantifying the relationship between respondent characteristics and interest in using the border-crossing approach, first using the baseline data only and subsequently using the follow-up data. When building the multivariable models, we first excluded variables that were highly correlated and then included variables in the final model if they modified the beta estimate for the primary predictor variable of interest, medication insurance status, by more than 10%. To investigate the possibility of falsified responses, we conducted a sensitivity analysis comparing the demographic characteristics and primary outcome responses among the decile of respondents with the shortest survey completion times with those of the total sample. In addition, we conducted a log file analysis to confirm that responses corresponded with unique internet protocol addresses. All quantitative analysis was done using R software (R Core Team).

### Sample Size Considerations

The sample size for this study was based on the minimum required sample size to ascertain, among individuals accessing the website, the proportion indicating an intention to use the border-crossing strategy with reasonable precision. There were no previous data to our knowledge on the likely value of this proportion, so we conservatively assumed that the true proportion was 0.5 (the value that generates the largest potential sample size requirements). We determined that a target sample size of 150 respondents for the baseline questionnaire would allow us to estimate the true prevalence with reasonable precision (plus or minus 8%) [16].

### Ethical Approval

This study was approved by the University of Toronto’s HIV Research Ethics Board and the St Michael’s Hospital Research Ethics Board before any study activities were initiated. All the participants documented their consent online before beginning either of the surveys.

## Results

### Survey Response

All 163 individuals who provided a valid email address were sent the baseline survey. Of the 158 participants who then began the baseline survey, 17 were excluded because of incomplete responses (n=5), repeat responses (n=6), and being non-Ontarian (n=6). Of the 141 participants included in the final baseline analysis, 110 provided complete responses to the follow-up survey and were included in the final follow-up analysis. The number of unique site visitors (ie, participants exposed to the survey link) could not be calculated as the survey link was likely shared through multiple informal channels.

### Demographics

Participant characteristics are described in Table 1. Median age was 23 (interquartile range [IQR] 22-25) years at baseline, and none of the respondents were below 18 years old. Most of the baseline participants were black (79.4%; 112/141), gay (87.2%; 123/141), and male (98.6%; 139/141), and 62.4% (88/141) had a college or undergraduate education or higher. Less than half (45.4%; 64/141) had drug coverage but most (83.0%; 117/141) had a family doctor or nurse practitioner with whom they felt comfortable discussing their sexual health. An HIRI-MSM score could be calculated for the 139 participants who answered all the requisite questions. The median HIRI-MSM score was 29 (IQR 26-30), and 97.8% (136/139) of men scored more than or equal to 10, meeting the index's criteria for high HIV risk. Characteristics of the 110 participants who completed the follow-up survey and the 31 who were lost to follow-up were similar (data not shown). None of the demographic or behavioral variables were significantly associated with completion of the follow-up survey.

**Table 1.** Participant characteristics.

| Characteristic | Baseline (n=141) | Follow-up (n=110) |
|---|---|---|
| Age (years), median (IQR[a]) | 23 (22-25) | 23 (22-25) |
| **Ethnicity, n (%)** | | |
| Black | 112 (79) | 88 (80) |
| White | 16 (11) | 13 (12) |
| Southeast Asian | 5 (4) | 4 (4) |
| Latin American | 3 (2) | 2 (2) |
| East Asian | 3 (2) | 1 (1) |
| South Asian | 2 (1) | 2 (2) |
| **Education, n (%)** | | |
| High school diploma or less | 53 (38) | 38 (35) |
| College/undergraduate degree | 81 (57) | 65 (59) |
| Professional or graduate degree | 7 (5) | 7 (6) |
| **Total family income (Can $[b]), n (%)** | | |
| 0-29,999 | 28 (20) | 18 (16) |
| 30,000-59,000 | 76 (54) | 64 (58) |
| >60,000 | 35 (25) | 26 (24) |
| **Medication payment[c], n (%)** | | |
| Private insurance | 56 (40) | 43 (39) |
| Out of pocket | 77 (55) | 63 (57) |
| Government drug benefit | 8 (6) | 4 (4) |
| Has a primary care provider with whom they feel comfortable discussing sexual health, n (%) | 117 (83) | 103 (94) |
| Has previously used pre-exposure prophylaxis, n (%) | 37 (26) | 35 (32) |
| **Drug use in the past 6 months, n (%)** | | |
| Alcohol | 116 (82) | 105 (95) |
| Marijuana (weed) | 47 (33) | 42 (38) |
| Cocaine | 23 (16) | 24 (22) |
| Poppers (amyl nitrate) | 22 (16) | 12 (11) |
| Methamphetamines (crystal and speed) | 9 (6) | 4 (4) |
| Injectable drugs | 3 (2) | 2 (2) |
| **Lifetime diagnosis of a sexually transmitted infection, n (%)** | | |
| Genital herpes | 23 (16) | 10 (9) |
| Gonorrhea | 12 (9) | 7 (6) |
| Chlamydia | 8 (6) | 3 (3) |
| Genital or anal warts | 6 (4) | 3 (3) |
| Syphilis | 3 (2) | 4 (4) |
| Ever participated in chemsex/party and play[d], n (%) | 50 (35) | 43 (39) |
| High Incidence Risk Index for men who have sex with men, median (IQR)[e] | 29 (26-30) | 29 (26-32) |
| Number of male sex partners in past 3 months, median (IQR) | 5 (4-7) | 5 (3-7) |
| Number of times having condomless receptive anal sex in past 3 months, median (IQR) | 6 (3-14) | 8 (5-11) |
| Number of partners known to be HIV positive in past 3 months, median (IQR) | 0 (0-1) | 0 (0-1) |

[a]IQR: interquartile range.

[b]Two participants responded *Don't know* in both baseline and follow-up.

[c]Two participants did not respond in the baseline.

[d]One participant did not respond in the baseline.

[e]Two participants did not answer all of the requisite risk questions in the baseline survey.

## Primary Outcomes

In the primary analysis, 80.1% of participants indicated that they were likely to use the border-crossing approach. This proportion was similar at follow-up (79.1%; 87/110). More than half of the participants had first heard about the approach from a friend (55.3%; 78/141), but participant responses varied by ethnicity: the leading response among black participants was *from a friend* (66.1%; 74/112), compared with only 20.6%; 6/29) of nonblack respondents, and the leading response among nonblack participants was *media or online* (48.5%, 14/29).

At baseline, more than half of the participants had both decided that they wanted PrEP (67.4%, 95/141) and already discussed PrEP with a health care provider (62.4%, 88/141). However, only 46.1% (65/141) had found a health care provider who would be willing to prescribe PrEP and only 24.1% (34/141) had already obtained a prescription. Few participants had completed the steps of the cascade specific to border crossing such as obtaining a mailbox in the United States (Figure 1). At the 3-month follow-up, the 110 participants who had completed both surveys showed increased completion of preliminary steps but decreased completion of the final steps related to border crossing (Figure 1). Notably, only 10% (11/110) of follow-up respondents had decided to obtain PrEP with a different method such as purchasing from a Canadian pharmacy.

We explored the variables associated with our primary outcome (likelihood of using the border-crossing approach) in exploratory logistic regression analyses separately for the baseline and follow-up data (Table 2). At baseline, factors associated with likeliness to use the border-crossing approach in both univariable and multivariable models included being black (adjusted odds ratio [aOR] 5.73, 95% CI 2.06-16.61), paying for medications out of pocket (aOR 5.18, 95% CI 1.82-17.04), and having a provider who was thought to be willing to prescribe PrEP (aOR 4.42, 95% CI 1.63-12.41). At follow-up, the findings were generally similar, as shown in Table 2. Only 9 individuals changed from being likely to unlikely to use the border-crossing approach between baseline and follow-up, whereas 8 did the reverse.

**Figure 1.** Steps completed in the border crossing for pre-exposure prophylaxis cascade at baseline and 3-months follow-up (n=110 participants who completed both surveys). PrEP: pre-exposure prophylaxis.

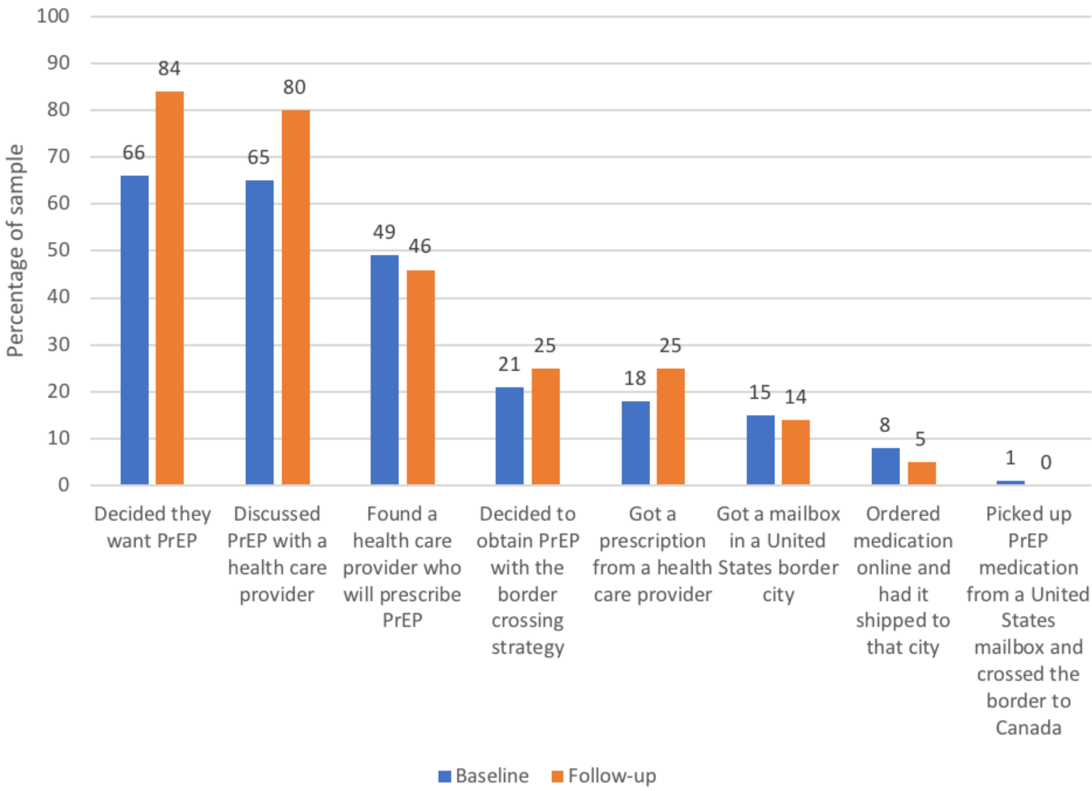


**Table 2.** Association between participant characteristics and the likelihood of using the border-crossing approach.

| Participant characteristic | Baseline | | | | Follow-up | | | |
|---|---|---|---|---|---|---|---|---|
| | Univariable | | Multivariable | | Univariable | | Multivariable | |
| | OR[a] (95% CI) | *P* value | OR (95% CI) | *P* value | OR (95% CI) | *P* value | OR (95% CI) | *P* value |
| **Demographics** | | | | | | | | |
| Age (years) | 0.74 (0.61-0.87) | .001 | —[b] | — | 0.86 (0.73-0.97) | .03 | — | — |
| **Ethnicity (ref: nonblack[c])** | | | | | | | | |
| Black | 8.20 (3.30-21.19) | <.001 | 5.73 (2.06-16.61) | .002 | 15.36 | <.001 | 14.42 (4.48-53.11) | <.001 |
| **Education (ref: high school diploma or less)** | | | | | | | | |
| College/undergraduate degree or more | 0.49 (0.18-1.19) | .10 | — | — | 0.79 (0.28-2.07) | .64 | — | — |
| **Total family income (Can $; ref: 0-29,999)** | | | | | | | | |
| 30,000-59,000 | 1.80 (0.60-5.14) | .29 | — | — | 3.49 (0.99-12.21) | .05 | — | — |
| >60,000 | 0.76 (0.24-2.31) | .62 | — | — | 0.69 (0.19-2.34) | .69 | — | — |
| **PrEP[d] access** | | | | | | | | |
| **Medication insurance (ref: private insurance or full government coverage)** | | | | | | | | |
| Out of pocket | 5.75 (2.26-16.72) | <.001 | 5.18 (1.82-17.04) | .003 | 7.20 (2.59-23.60) | <.001 | 6.69 (2.07-25.91) | .003 |
| Has a primary care provider that the respondent thinks would be willing to prescribe PrEP | 6.39 (2.66-16.00) | <.001 | 4.42 (1.63-12.41) | .004 | 5.77 (2.18-15.91) | <.001 | — | — |
| Has previously used PrEP | 5.83 (1.62-37.45) | .02 | — | — | 3.88 (1.21-17.37) | .04 | — | — |
| Able and willing to pay more than Can $100 per month for PrEP | 0.66 (0.28-1.53) | .31 | — | — | 1.02 (0.37-3.10) | .97 | — | — |
| **Risk factors** | | | | | | | | |
| High Incidence Risk Index for men who have sex with men score | 1.09 (1.03-1.15) | .002 | — | — | 1.10 (1.04-1.18) | .002 | — | — |
| Number of male sex partners in the past 3 months | 1.05 (0.90-1.23) | .63 | — | — | 1.03 (0.87-1.23) | .76 | — | — |
| **Perceived chance of getting HIV in the next year (ref: little to none)** | | | | | | | | |
| More than a little or greater | 0.97 (0.37-2.87) | .91 | — | — | 0.53 (0.20-1.42) | .19 | — | — |
| Ever diagnosed with a sexually transmitted infection | 1.94 (0.72-6.18) | .19 | — | — | 2.38 (0.80-8.76) | .15 | — | — |
| Party drug use in the past 6 months[e] | 1.59 (0.67-4.13) | .31 | — | — | 1.61 (0.62-4.56) | .34 | — | — |
| Ever participated in chemsex/party and play | 2.34 (0.93-6.76) | .12 | — | — | — | .16 | | — |
| **Perceptions of the border-crossing approach** | | | | | | | | |
| **First heard about the approach from (ref: friend[s])** | | | | | | | | |
| Other (health care provider, community organization, online, etc) | 0.25 (0.10-0.60) | .003 | — | — | 0.20 (0.06-0.52) | .002 | — | — |

| Participant characteristic | Baseline | | | | Follow-up | | | |
|---|---|---|---|---|---|---|---|---|
| | Univariable | | Multivariable | | Univariable | | Multivariable | |
| | OR[a] (95% CI) | *P* value | OR (95% CI) | *P* value | OR (95% CI) | *P* value | OR (95% CI) | *P* value |
| Number of barriers perceived/anticipated | 1.19 (0.88-1.65) | .28 | — | — | 0.97 (0.73-1.31) | .87 | — | — |
| Concerned about the legality of the approach | 1.47 (0.63-3.44) | .40 | — | — | 1.67 (0.66-4.39) | .28 | — | — |
| Knows someone who uses the approach | 6.70 (2.53-21.20) | <.001 | — | — | 2.42 (0.95-6.97) | .07 | — | — |

[a]OR: odds ratio.

[b]Further analysis not run on these variables.

[c]White, East Asian, South Asian, and Latin American.

[d]PrEP: pre-exposure prophylaxis.

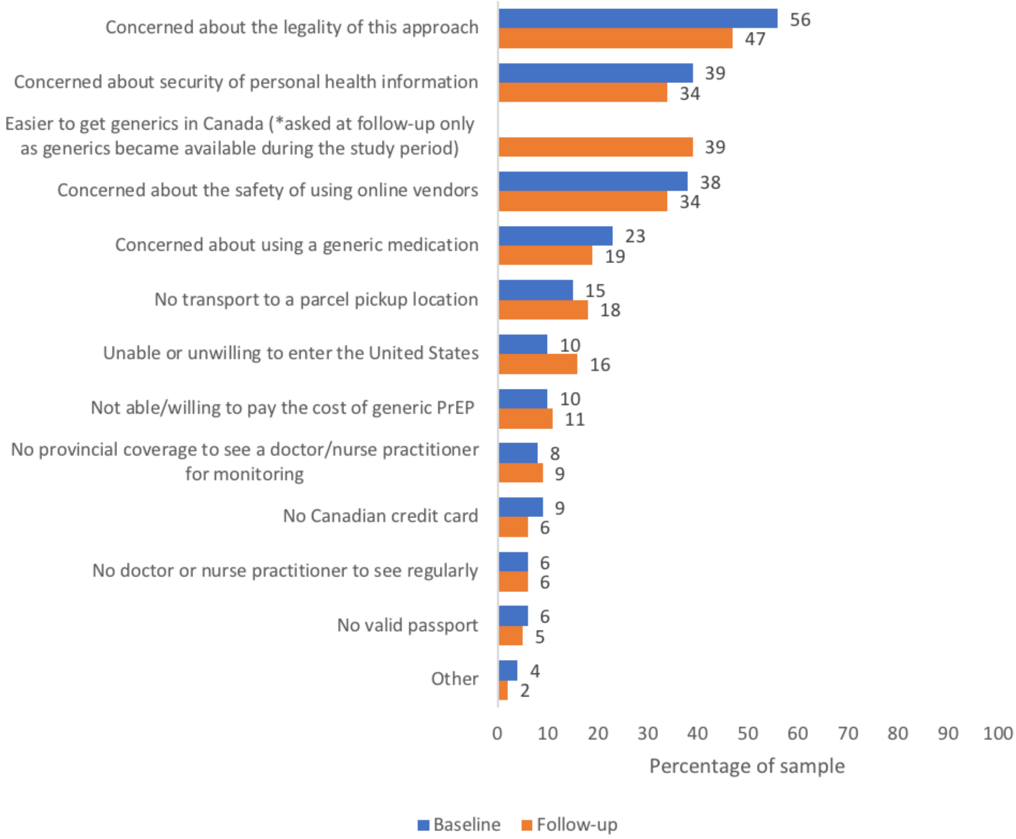
[e]Party drugs specified as methamphetamines, gamma-hydroxybutyrate, amyl nitrate, cocaine, crack, ketamine, and injectable drugs.

## Motivators and Barriers

The most commonly identified motivator to using the border-crossing approach was the high cost of PrEP at Canadian pharmacies (59.6%, 84/141), with 57.4% (81/141)of the participants identifying a maximum monthly amount that they would be able and willing to pay for PrEP that was less than Can $100. The second most commonly identified motivator was already knowing someone who uses the approach (52.5%, 74/141).

The most commonly identified barriers to using the border-crossing approach at baseline were concerns about the legality of the approach (56.0%, 79/141), the safety of using online vendors (38.3%, 54/141), and the security of personal health information (36.9%; 52/141; Figure 2). The findings for motivators and barriers were similar at follow-up, except that 40.9% (45/110) of participants identified a new barrier being that it had become easier to get generic PrEP in Canada instead. We added this response option for the follow-up questionnaire only because domestically manufactured generic PrEP became available and partially publicly subsidized in Ontario in late 2017, before follow-up data collection began. Although 44.5% (49/110) of follow-up participants reported that they would be less likely to use the border-crossing approach given these policy changes, 86.3% (95/110) were unaware of the new changes before beginning the follow-up survey.

**Figure 2.** Reported barriers to using the border-crossing strategy. PrEP: pre-exposure prophylaxis.




### Checking Response Quality

We conducted a sensitivity analysis comparing the decile of respondents (n=14) with the shortest survey completion times to the total sample. Response proportions for both groups were consistent across both the primary outcome and the demographic characteristics included in Table 1 (data not shown).

## Discussion

### Principal Findings

Understanding both who is considering online alternatives for PrEP access and why they are doing so is vital to understanding gaps in PrEP implementation. In our sample of mostly Toronto-based, black GBMSM, we found high interest in pursuing an innovative approach to obtaining PrEP medications, involving online ordering and cross-border retrieval, measuring 80.1% (113/141) at baseline and 79.1% (87/110) at follow-up. Participants identified both key motivators (eg, knowing someone who uses a border-crossing approach and high domestic costs) and barriers (eg, concerns about the legality of such an approach, security of personal health information, and the safety of online vendors) to pursuing this strategy. To our knowledge, this is the first study regarding this Web-based, border-crossing approach to obtain PrEP medications.

The proportion of black participants was surprisingly high at 79.0% (112/141) , given that less than 5% of Ontario's men identify as black [17], raising the possibility of sampling bias. However, the direction of any such bias is unclear, and our primary purpose was to assess the intentions of those accessing the online resource rather than those of the GBMSM population as a whole. We suspect that the high number of black participants is a result of social networks circulating the online link to the study survey; 66.1% (74/112) of black participants reported that they had heard of the border-crossing strategy from friends vs only 22.9% (6/29) of nonblack participants. Given that North American black GBMSM have historically been both socioeconomically disadvantaged and harder to engage in HIV prevention research [18-21], our high number of black respondents may reflect an unmet need for alternative methods of PrEP access in this population, and the advantages of using Web-based approaches for reaching and studying the needs of marginalized populations.

### Comparison With Previous Work

Online PrEP shopping has been a major alternative approach for accessing PrEP among GBMSM in other jurisdictions such as the United Kingdom and Australia. Online purchasing of inexpensive PrEP likely played a causal role in lowering the rates of HIV infection among GBMSM at sexual health clinics in the United Kingdom [22,23] and filled a critical access gap in the formal health care system, where the National Health Service has decided not to fund PrEP medication [24]. UK online information pages such as *I want PrEP Now* were precursors to the Web pages created by Canadian activists and community organizations (eg, *The Davie Buyers Club* and the GMSH's *The PrEP You Want* campaign) [25], and many of these pages refer users to the same pool of Web-based international generic medication suppliers. Similarly in Australia, online purchasing, often championed by community groups and websites, was the primary means of PrEP access before the eventual public listing of PrEP on the national drug benefit scheme in 2018 [26,27]. In British Columbia, limited data from 2017 suggest that 200 individuals accessed PrEP through *The Davie Buyers Club* over a 3-month period, though detailed information on the characteristics and experiences of these individuals is lacking [9]. Although there has been some concern over the quality of medications ordered online, both in our sample and in the literature [28,29], evidence from both the UK and Canadian contexts suggests that online PrEP distributors provide safe and effective generic medications [30,31]. However, given the complexity and inconvenience involved in border crossing to obtain PrEP, we contend that this strategy is not the optimal way to make this primary prevention strategy available in Canada, and PrEP needs to be more widely available to Canadians at risk.

Following in the mold of HIV cascades, hypothetical PrEP cascades have been proposed and used to identify bottlenecks at different stages of PrEP implementation [32-34]. For instance, our previous hypothetical PrEP cascade found that although 64.4% of Toronto GBMSM undergoing anonymous HIV testing were at objectively high HIV risk, 68.3% did not perceive themselves to be at elevated risk and 47.6% lacked access to a family doctor [7]. Our hypothetical cascade of border-crossing steps identified parallel bottlenecks in accessing health care providers and obtaining PrEP prescriptions. These data highlight the need for continued efforts to make PrEP providers and PrEP medication widely accessible, even in large urban centers such as Toronto.

Two positive developments in Ontario's PrEP access landscape have been the entry of generics into the Canadian PrEP market (August 2017) and the government's initiation of partial public subsidization (September 2017). As a result, PrEP costs have been reduced (the generic currently costs approximately Can $220 per month) or largely covered (for Ontarians meeting the criteria for public drug benefits, sometimes involving copayment) [35]. As indicated by the 40.9% (45/110) of follow-up study participants who cited the availability of cheap generics in Canada as a deterrent to pursuing a border-crossing approach, this shift in cost has made conventional approaches to obtaining PrEP more viable for many potential users. However, despite the availability of reduced-cost generics in Ontario, our previous work demonstrated that PrEP could in some cases be obtained online for as little as Can $33 per month [31], representing a cost savings of approximately $187 per month over the generic TDF/FTC available in Ontario.

However, PrEP remains financially inaccessible for many potential users in Canada. This held true in our sample where 57.4% (81/141) participants identified a maximum monthly amount that they would be able and willing to pay for PrEP that was less than Can $100. Although Can $33 per month for online PrEP is within this range, the attrition we observed in our border-crossing cascade suggests that relying on online purchasing, even at this price, is not a viable public health strategy for improving PrEP accessibility. Given the recent findings of real-world reductions in HIV incidence of up to 100% in PrEP users, and PrEP's role in the substantial reductions in HIV incidence in settings such as San Francisco and New South Wales [36-39], Canada's inconsistent access to this major public health tool emphasizes the urgent need for truly universal coverage. This is particularly important given the failure of HIV incidence to decline among Canadian GBMSM in recent years [40].

## Limitations

This study has several limitations that warrant consideration. First, because our study was an anonymous open online survey, it was not possible to exclude repeat participants, and some may have been motivated by the modest financial compensation. We attempted to minimize repeats by requiring participants to enter unique study codes that were assigned to each respondent's email address. We also checked for possible falsified responses by conducting a sensitivity analysis, which showed consistency between the demographic characteristics and primary outcome responses among the fastest decile of responders and the total sample. Second, hypothetical bias may have inflated the number of participants indicating likeliness to use this fairly time-intensive border-crossing approach [41]. We addressed this possibility in part by asking about the actual completion of cascade steps at 3-month follow-up. In addition, however, the relatively short time frame (3 months) between baseline and follow-up may still have been insufficient to capture participants' progress in using the border-crossing approach. These last two biases may in part explain why we found no participants who had completed every step of the border-crossing approach at follow-up. Furthermore, it is notable that although half of the respondents reportedly knew someone using the border-crossing approach, very few respondents actually did so themselves. This discrepancy raises the possibilities that respondents were underreporting their use of the procedure, that our study underrecruited individuals already successfully using this strategy, or that a limited number of border crossers were known to a large number of respondents. Finally, because we calculated HIRI-MSM scores for our participants by doubling the responses given for the past 3 months (rather than directly asking about the past 6 months), they may have deviated slightly from *true* scores. Moreover, HIRI-MSM scoring may be less relevant to the contemporary context as HIV risk through condomless sex can be reduced to zero with HIV treatment as prevention and/or PrEP. Regardless, our sample was likely still at high HIV risk because of its young age, high mean number of sex partners, frequent condomless receptive anal sex, and a 35% participation rate in chemsex.

## Conclusions

PrEP is a promising tool for curbing high rates of HIV in GBMSM communities, but access remains highly variable in Canada. Our results suggest that despite the falling price of PrEP medications and some government subsidization, many potential PrEP users remain interested in alternative, cheaper methods of obtaining PrEP medications such as online shopping and border crossing. Effective PrEP implementation will require alternative strategies such as universal public coverage to ensure readily accessible PrEP medications for GBMSM living in Canada.

## Acknowledgments

The authors would like to acknowledge the survey respondents, the Gay Men's Sexual Health Alliance, Black CAP, and St. Michael's Hospital for their input and support.

This work was supported by a grant from the Canadian Institutes of Health Research (grant number HIM 145368). DHST is supported by a New Investigator Award from the Canadian Institutes of Health Research and the Ontario HIV Treatment Network. AS is the creator of the Davie Buyers Club resource and DG is a cocreator of The PrEP You Want resource.

## Conflicts of Interest

In the past 3 years, DHST has received research grants from ViiV Healthcare and Gilead Sciences and has been a site principal investigator for clinical trials sponsored by GlaxoSmithKline.

## References

1. Bourgeois AC, Edmunds M, Awan A, Jonah L, Varsaneux O, Siu W. Government of Canada. 2017. HIV in Canada—Supplementary Tables, 2016 URL: https://www.canada.ca/en/public-health/services/reports-publications/canada-communicable-disease-report-ccdr/monthly-issue/2017-43/ccdr-volume-43-12-december-7-2017/hiv-2016-supplementary-tables.html [WebCite Cache ID 74ENfi3cu]
2. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
3. Anderson PL, Glidden DV, Liu A, Buchbinder S, Lama JR, Guanira JV, iPrEx Study Team. Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. Sci Transl Med 2012 Sep 12;4(151):151ra125 [FREE Full text] [doi: 10.1126/scitranslmed.3004006] [Medline: 22972843]
4. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, iPrEx Study Team. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829 [FREE Full text] [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
5. Molina JM, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, ANRS IPERGAY Study Group. On-demand preexposure prophylaxis in men at high risk for HIV-1 infection. N Engl J Med 2015 Dec 3;373(23):2237-2246. [doi: 10.1056/NEJMoa1506273] [Medline: 26624850]
6. Wilton J, Kain T, Fowler S, Hart TA, Grennan T, Maxwell J, et al. Use of an HIV-risk screening tool to identify optimal candidates for PrEP scale-up among men who have sex with men in Toronto, Canada: disconnect between objective and subjective HIV risk. J Int AIDS Soc 2016;19(1):20777 [FREE Full text] [doi: 10.7448/IAS.19.1.20777] [Medline: 27265490]
7. Formulary Search. URL: https://www.formulary.health.gov.on.ca/formulary/ [WebCite Cache ID 74ENpmKEE]
8. Immunodeficiency Clinic. 2018. Provincial/Territorial Coverage of ARV Drugs for HIV Prevention Across Canada:Post-exposure Prophylaxis (PEP) and Pre-exposure Prophylaxis (PrEP) URL: https://hivclinic.ca/wp-content/uploads/2018/04/ARV-Coverage_PEP-and-PrEP.pdf [WebCite Cache ID 74ENzf1xa]
9. Smith A. Davie Buyers club: history and current state of the generic PrEP importation pathway. In: Proceedings of the 13th Annual Gay Men's Health Summit. 2017 Presented at: Romancing the Package'13; November 2-4, 2017; Vancouver, Canada URL: https://d3n8a8pro7vhmx.cloudfront.net/vancitystudios/pages/19322/attachments/original/1534990785/cbrc_summit17_web-2_0.pdf?1534990785
10. Government of Canada: Justice Laws. 2018. Food and Drug Regulations (CRC, c 870) URL: http://laws.justice.gc.ca/eng/regulations/c.r.c.,_c._870/index.html [WebCite Cache ID 74EO9aD0t]
11. Davie Buyers Club. 2018. How to Access PrEP Without Insurance in Canada (for $45 CAD Per Month) URL: https://daviebuyersclub.wordpress.com/ [WebCite Cache ID 74EOUMD7c]
12. The Sex You Want. 2016. The PrEP You Want: How to Order PrEP Online URL: http://thesexyouwant.ca/wp-content/uploads/2016/06/The-PrEP-You-Want-how-to-order-PrEP-Online-final-updated-1.pdf [WebCite Cache ID 74EOed40B]
13. Rana J, Wilton J, Fowler S, Hart TA, Bayoumi AM, Tan DH. Trends in the awareness, acceptability, and usage of HIV pre-exposure prophylaxis among at-risk men who have sex with men in Toronto. Can J Public Health 2018 Dec;109(3):342-352 [FREE Full text] [doi: 10.17269/s41997-018-0064-3] [Medline: 29981088]
14. Wray R, Branka A, Bennett-AbuAyyash C, Tuck A. St. Michael's Hospital. 2013. We Ask Because We Care: The Tri-Hospital + TPH Health Equity Data Collection Research Project Report URL: http://www.stmichaelshospital.com/quality/equity-data-collection-report.pdf [WebCite Cache ID 74EOigtzM]
15. Smith DK, Pals SL, Herbst JH, Shinde S, Carey JW. Development of a clinical screening index predictive of incident HIV infection among men who have sex with men in the United States. J Acquir Immune Defic Syndr 2012 Aug 1;60(4):421-427. [doi: 10.1097/QAI.0b013e318256b2f6] [Medline: 22487585]
16. Naing L, Winn T, Rusli BN. CiteSeerX. 2006. Practical Issues in Calculating the Sample Size for Prevalence Studies URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.504.2129&rep=rep1&type=pdf

17. Statistics Canada. 2016. Census Profile, 2016 Census URL: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E [WebCite Cache ID 74EOn33YS]
18. Millett GA, Peterson JL, Flores SA, Hart TA, Jeffries WL, Wilson PA, et al. Comparisons of disparities and risks of HIV infection in black and other men who have sex with men in Canada, UK, and USA: a meta-analysis. Lancet 2012 Jul 28;380(9839):341-348. [doi: 10.1016/S0140-6736(12)60899-X] [Medline: 22819656]
19. Eaton LA, Driffin DD, Bauermeister J, Smith H, Conway-Washington C. Minimal awareness and stalled uptake of pre-exposure prophylaxis (PrEP) among at risk, HIV-negative, black men who have sex with men. AIDS Patient Care STDS 2015 Aug;29(8):423-429 [FREE Full text] [doi: 10.1089/apc.2014.0303] [Medline: 26083143]
20. Smith DK, Toledo L, Smith DJ, Adams MA, Rothenberg R. Attitudes and program preferences of African-American urban young adults about pre-exposure prophylaxis (PrEP). AIDS Educ Prev 2012 Oct;24(5):408-421. [doi: 10.1521/aeap.2012.24.5.408] [Medline: 23016502]
21. Matthews DD, Herrick AL, Coulter RW, Friedman MR, Mills TC, Eaton LA, POWER Study Team. Running backwards: consequences of current HIV incidence rates for the next generation of black MSM in the United States. AIDS Behav 2016 Jan;20(1):7-16 [FREE Full text] [doi: 10.1007/s10461-015-1158-z] [Medline: 26267251]
22. Brown AE, Mohammed H, Ogaz D, Kirwan PD, Yung M, Nash SG, et al. Fall in new HIV diagnoses among men who have sex with men (MSM) at selected London sexual health clinics since early 2015: testing or treatment or pre-exposure prophylaxis (PrEP)? Euro Surveill 2017 Jun 22;22(25):pii: 30553 [FREE Full text] [doi: 10.2807/1560-7917.ES.2017.22.25.30553] [Medline: 28662762]
23. Aloysius I, Savage A, Zdravkov J, Korologou-Linden R, Hill A, Smith R, et al. InterPrEP. Internet-based pre-exposure prophylaxis with generic tenofovir DF/emtricitabine in London: an analysis of outcomes in 641 patients. J Virus Erad 2017 Oct 1;3(4):218-222 [FREE Full text] [doi: 10.1111/hiv.12528] [Medline: 29057086]
24. Kirby T. HIV pre-exposure prophylaxis: a tale of two countries. Lancet Infect Dis 2017 Dec;17(1):32-33. [doi: 10.1016/S1473-3099(16)30562-X] [Medline: 27998575]
25. I Want PrEP Now. URL: https://www.iwantprepnow.co.uk/ [WebCite Cache ID 74EOwHIYa]
26. Holt M, Lea T, Schmidt H, Kolstee J, Ellard J, Murphy D, et al. Willingness to use and have sex with men taking HIV pre-exposure prophylaxis (PrEP): results of online surveys of Australian gay and bisexual men, 2011-2015. Sex Transm Infect 2017 Dec;93(6):438-444. [doi: 10.1136/sextrans-2016-052774] [Medline: 28108701]
27. Holt M, Lea T, Bear B, Halliday D, Ellard J, Murphy D, et al. Trends in attitudes to and the use of HIV pre-exposure prophylaxis by Australian gay and bisexual men, 2011-2017: implications for further implementation from a diffusion of innovations perspective. AIDS Behav 2019 Jul;23(7):1939-1950. [doi: 10.1007/s10461-018-2368-y] [Medline: 30539496]
28. Baert B, de Spiegeleer B. Quality analytics of internet pharmaceuticals. Anal Bioanal Chem 2010 Sep;398(1):125-136. [doi: 10.1007/s00216-010-3912-4] [Medline: 20582403]
29. Wang T, Hoag SW, Eng ML, Polli J, Pandit NS. Quality of antiretroviral and opportunistic infection medications dispensed from developing countries and internet pharmacies. J Clin Pharm Ther 2015 Feb;40(1):68-75. [doi: 10.1111/jcpt.12226] [Medline: 25381836]
30. Wang X, Nwokolo N, Korologou-Linden R, Hill A, Whitlock G, Day-Weber I, et al. InterPrEP: internet-based pre-exposure prophylaxis with generic tenofovir disoproxil fumarate/emtrictabine in London - analysis of pharmacokinetics, safety and outcomes. HIV Med 2018 Dec;19(1):1-6. [doi: 10.1111/hiv.12528] [Medline: 28657199]
31. Lott A, Naccarato M, Turton R, Tan DH. POPPEE: Purchasing Online Pre-Exposure Prophylaxis (PrEP) Pills to Evaluate Equivalence. In: Proceedings of the 26th Annual Canadian Conference on HIV/AIDS Research. 2017 Presented at: CAHR'17; April 6-9, 2017; Montreal, Quebec.
32. Liu A, Colfax G, Cohen S, Bacon O, Kolber M, Amico KR, et al. The Spectrum of Engagement in HIV Prevention: Proposal for a PrEP Cascade. In: Proceedings of the 7th International Conference on HIV Treatment and Prevention Adherence. 2012 Presented at: Adherence'12; June 3-5, 2012; London, UK.
33. Parsons JT, Rendina HJ, Lassiter JM, Whitfield TH, Starks TJ, Grov C. Uptake of HIV pre-exposure prophylaxis (PrEP) in a national cohort of gay and bisexual men in the United States. J Acquir Immune Defic Syndr 2017 Dec 1;74(3):285-292 [FREE Full text] [doi: 10.1097/QAI.0000000000001251] [Medline: 28187084]
34. Kelley CF, Kahle E, Siegler A, Sanchez T, del Rio C, Sullivan PS, et al. Applying a PrEP continuum of care for men who have sex with men in Atlanta, Georgia. Clin Infect Dis 2015 Nov 15;61(10):1590-1597 [FREE Full text] [doi: 10.1093/cid/civ664] [Medline: 26270691]
35. Ontario. 2018. Get Coverage for Prescription Drugs URL: https://www.ontario.ca/page/get-coverage-prescription-drugs [WebCite Cache ID 74EQ6D7SU]
36. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet 2016 Jan 2;387(10013):53-60 [FREE Full text] [doi: 10.1016/S0140-6736(15)00056-2] [Medline: 26364263]
37. Volk JE, Marcus JL, Phengrasamy T, Blechinger D, Nguyen DP, Follansbee S, et al. No new HIV infections with increasing use of HIV preexposure prophylaxis in a clinical practice setting. Clin Infect Dis 2015 Nov 15;61(10):1601-1603 [FREE Full text] [doi: 10.1093/cid/civ778] [Medline: 26334052]

38. San Francisco Department of Public Health. 2016. HIV Epidemiology: Annual Report 2016 URL: https://www.sfdph.org/dph/files/reports/RptsHIVAIDS/Annual-Report-2016-20170831.pdf [WebCite Cache ID 74EQDuSgt]
39. Grulich AE, Guy R, Amin J, Jin F, Selvey C, Holden J, Expanded PrEP Implementation in Communities New South Wales (EPIC-NSW) research group. Population-level effectiveness of rapid, targeted, high-coverage roll-out of HIV pre-exposure prophylaxis in men who have sex with men: the EPIC-NSW prospective cohort study. Lancet HIV 2018 Dec;5(11):e629-e637. [doi: 10.1016/S2352-3018(18)30215-7] [Medline: 30343026]
40. The Ontario HIV Epidemiology and Surveillance Initiative. 2016. New HIV Diagnoses in Ontario: Preliminary Update, 2016 URL: http://www.ohesi.ca/documents/OHESI-New-HIV-Diagnoses-preliminary-updates.pdf [WebCite Cache ID 76cCFxayD]
41. List JA, Gallet CA. What experimental protocol influence disparities between actual and hypothetical stated values? Environ Resour Econ 2001;20(3):241-254. [doi: 10.1023/A:1012791822804]


## Abbreviations

**aOR:** adjusted odds ratio
**GBMSM:** gay, bisexual, and other men who have sex with men
**GMSH:** Gay Men’s Sexual Health Alliance
**HIRI-MSM:** High Incidence Risk Index for men who have sex with men
**IQR:** interquartile range
**PrEP:** pre-exposure prophylaxis
**TDF/FTC:** tenofovir disoproxil fumarate and emtricitabine


*Edited by G Eysenbach; submitted 28.11.18; peer-reviewed by M Holt, T Starks; comments to author 10.01.19; revised version received 07.03.19; accepted 16.04.19; published 22.07.19.*

*Please cite as:*
*Walmsley B, Gallant D, Naccarato M, Hull M, Smith A, Tan DHS*
*The PrEP You Want: A Web-Based Survey of Online Cross-Border Shopping for HIV Prophylaxis Medications*
*J Med Internet Res 2019;21(7):e12076*
*URL: http://www.jmir.org/2019/7/e12076/*
*doi:10.2196/12076*
*PMID:31333200*

Review

# Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review

Syed Jamal Safdar Gardezi[1], PhD; Ahmed Elazab[1], PhD; Baiying Lei[1], PhD; Tianfu Wang[1], PhD

National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong, Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

**Corresponding Author:**
Baiying Lei, PhD
National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong
Key Laboratory for Biomedical Measurements and Ultrasound Imaging
School of Biomedical Engineering, Health Science Center, Shenzhen University
Shenzhen,
China
Phone: 86 13418964616
Fax: 86 0755 86172219
Email: leiby@szu.edu.cn

## Abstract

**Background:** Machine learning (ML) has become a vital part of medical imaging research. ML methods have evolved over the years from manual seeded inputs to automatic initializations. The advancements in the field of ML have led to more intelligent and self-reliant computer-aided diagnosis (CAD) systems, as the learning ability of ML methods has been constantly improving. More and more automated methods are emerging with deep feature learning and representations. Recent advancements of ML with deeper and extensive representation approaches, commonly known as deep learning (DL) approaches, have made a very significant impact on improving the diagnostics capabilities of the CAD systems.

**Objective:** This review aimed to survey both traditional ML and DL literature with particular application for breast cancer diagnosis. The review also provided a brief insight into some well-known DL networks.

**Methods:** In this paper, we present an overview of ML and DL techniques with particular application for breast cancer. Specifically, we search the PubMed, Google Scholar, MEDLINE, ScienceDirect, Springer, and Web of Science databases and retrieve the studies in DL for the past 5 years that have used multiview mammogram datasets.

**Results:** The analysis of traditional ML reveals the limited usage of the methods, whereas the DL methods have great potential for implementation in clinical analysis and improve the diagnostic capability of existing CAD systems.

**Conclusions:** From the literature, it can be found that heterogeneous breast densities make masses more challenging to detect and classify compared with calcifications. The traditional ML methods present confined approaches limited to either particular density type or datasets. Although the DL methods show promising improvements in breast cancer diagnosis, there are still issues of data scarcity and computational cost, which have been overcome to a significant extent by applying data augmentation and improved computational power of DL algorithms.

***(J Med Internet Res 2019;21(7):e14464)*** doi:10.2196/14464

**KEYWORDS**

breast cancer; lesion classification; malignant tumor; machine learning; convolutional neural networks; deep learning

## Introduction

Cancer is one of the leading causes of female deaths worldwide. It has caused more deaths than any other diseases such as tuberculosis or malaria. The World Health Organization (WHO) agencies for cancer research (ie, International agency for cancer research (IARC) and American Cancer Society) report that 17.1 million new cancer cases are recorded in 2018 worldwide [1]. WHO estimates that cancer incidences might increase to 27.5 million by 2040, with an estimated 16.3 million deaths expected as a result of cancer [1].

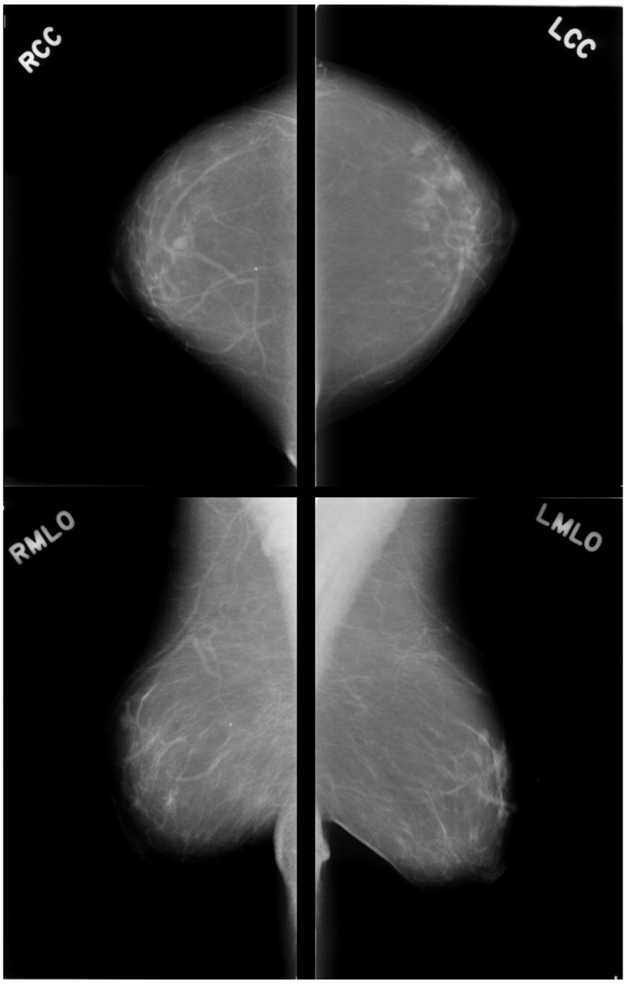
Breast cancer is among the 4 leading cancers in women worldwide (ie, lung, breast and bowel [including anus], stomach, and prostate cancers). The IARC statistics show that breast cancer accounts for 25% of all cancer cases diagnosed in women worldwide. Around 53% of these cases come from developing countries, which represent 82% of the world population [1]. It is reported that 626,700 deaths will occur only in 2018 [1]. Breast cancer is the leading cause of cancer death among women in developing countries and the second leading cause of cancer death (following lung cancer) among women in developed countries.

In breast, the cancer cells may spread to lymph nodes or even cause damage to other parts of the body such as lungs. Breast cancer more often starts from the malfunctioning of milk-producing ducts (invasive ductal carcinoma). However, it may also begin in the glandular tissues called lobules or other cells or tissues within the breast [1]. Researchers have also found that hormonal, lifestyle, and environmental changes also contribute to increasing the risk of breast cancer [2,3].

To visualize the internal breast structures, a low-dose x-ray of the breasts is performed; this procedure is known as mammography in medical terms. It is one of the most suitable techniques to detect breast cancer. Mammograms expose the breast to much lower doses of radiation compared with devices used in the past [4]. In recent years, it has proved to be one of the most reliable tools for screening and a key method for the early detection of breast cancer [5,6]. The mammograms are acquired at 2 different views for each breast: craniocaudal (CC) view and mediolateral oblique (MLO) view (Figure 1).

In this review, we present the recent work in breast cancer detection using conventional machine learning (ML) and deep learning (DL) techniques. The aim of this work was to provide the reader with an introduction to breast cancer literature and recent advancements in breast cancer diagnosis using multiview digital mammograms (DMs). The survey aimed to highlight the challenges in the application of DL for early detection of breast cancer using the multiview digital mammographic data. We present the recent studies that have addressed these challenges and finally provide some insights and discussions on the current open problems. This review is divided into 2 major parts. The first part presents a brief introduction of different steps of a conventional ML method (ie, enhancement, feature extraction, segmentation, and classification), whereas the second part focuses on DL techniques, with an emphasis on multiview (ie, CC and MLO) mammographic data. The present DL literature can be characterized for breast density discrimination, detection, and classification of the lesion in breast cancer in the multiview digital mammographic data. The rest of this review is organized as follows.

**Figure 1.** Multiview breast mammogram of a patient. The first column presents two views of the right breast: right craniocaudal (RCC) view and right mediolateral oblique (RMLO) view. The second column presents two views of the left breast: left craniocaudal (LCC) view and left mediolateral oblique (LMLO) view.



## Methods

### Conventional Machine Learning Pipeline

In this section, we present various steps involved in a computer-aided diagnosis (CAD) system using the conventional workflow. The steps involved are outlined in Figure 2 and are discussed briefly as follows.

### Breast Profiling and Preprocessing

Mammogram preprocessing is one of the primary steps in a CAD system. In the preprocessing step, the unwanted objects are removed from the mammograms, which include annotations, labels, and background noises as can be seen in Figure 3. The preprocessing helps the localization of region for abnormality search. In mammogram preprocessing, one of the major challenges is to accurately define the pectoral muscle (PM) boundary from the rest of the breast region. The PMs are mostly present in MLO views of the mammograms. The presence of PMs in the MLO view can interrupt the automatic detection of lesions and can increase the false positive (FP) alarms. Many studies advocated the removal of PMs [7-15] for improving the diagnostic accuracy of the CAD system. Thus, successful removal of PMs is vital to avoid false detection. Moreover, it also reduces the time complexity and improves the accuracy apart from avoiding the intra-observation discrepancies.

**Figure 2.** Difference between 2 pipelines: conventional machine learning pipeline (left) and deep learning pipeline (right).

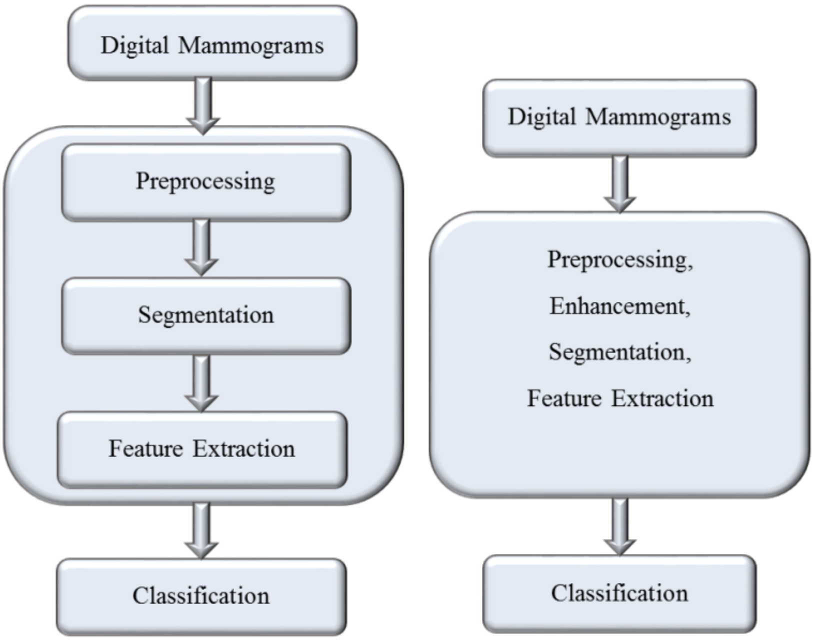
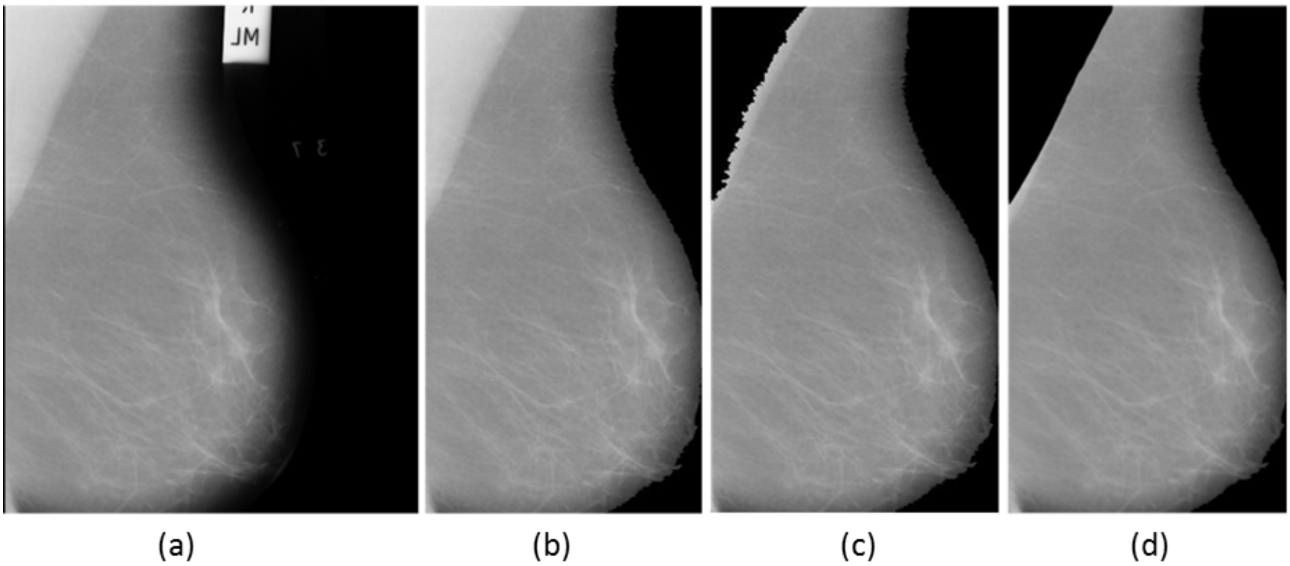



**Figure 3.** (a) Original mammogram image 1024×1024. (b) Preprocessing to remove annotations. (c) pectoral muscle (PM) removal by region growing. (d) PM removal by adaptive segmentation.



## Image Enhancement Techniques

Image enhancement techniques are used to improve the mammogram's quality in terms of improving the contrast and enhancing its readability. It helps the system to detect the mammographic lesions with poor visibility and contrast by improving it. The major goal of mammogram enhancement is to improve the image quality on the mammograms with low contrast. The low-contrast regions with small abnormalities are often concealed in surrounding tissues, leading to a misdiagnosis. The image enhancements improve the overall quality of the images, thus making it relatively easier for the reader and CAD systems to detect these subtle abnormalities. The enhancements may add distortions to the anatomical characteristics of an image or amplify the noises. Thus, only those methods would be acceptable that maintain a similar appearance to the original image. Recently, with the introduction of direct digital technology in mammography, with dynamic range, improved contrast, and signal to noise ratio, there is a limited scope of mammogram enhancement.

The enhancement techniques are generally divided into 3 categories: spatial domain, frequency domain, and a combination of spatial and frequency domain techniques [16]. However, these techniques can be characterized into 4 types [17] based on their particular usage: namely, conventional, region-based, feature-based, and fuzzy enhancement techniques. The primary aim of enhancements is to improve the quality of mammograms to achieve high diagnostic performance. The conventional methods can be adapted for local as well as global enhancement of the mammograms. However, the conventional methods have a tendency to enhance the noise factor as well. On the other hand, the region-based methods are suitable for contrast enhancements of particular regions of interest (ROIs) with varying shapes and sizes. The region-based methods help to enhance the anatomical details of the ROIs without any addition of artifacts. These methods are generally well suited for microcalcification enhancements in breasts with dense tissues. The feature-based enhancement techniques are applied on mammograms with calcifications as well as masses. The multiscale transforms such as wavelets are used because of their dilation and translation properties that are best suited for nonstationary signals. The low frequencies are suppressed, whereas only higher frequencies are kept by applying a threshold. Thus, the reconstructed images only contain highest frequencies with possible lesion regions. Finally, the fuzzy enhancement technique uses the maximum fuzzy entropy principle on the normalized mammograms to enhance the contrast and suppress the noise. These techniques are effective to enhance the mass contours and present the fine details of mammogram features.

## Mammographic Mass Segmentation Techniques

The segmented region is vital for feature extraction and detection of abnormal tissues in the breast, and it needs to be well focused and precise. Therefore, the segmentation is important to extract an ROI that provides a precise measurement of breast regions with abnormalities and normal regions. Segmentation involves the fundamental step of separating the breast region from the background and aims to separate the breast regions from the other objects. It is an important step to preserve the margin characteristics of mammograms before any further processing.

The segmentation aims to extract ROIs with possible masses, and it may involve partitioning of the mammogram into several nonoverlapping regions with candidate mass lesions. At the detection stages, higher sensitivity rate and more FPs are expected. Figure 4 illustrates the FP detection at pixel level compared with ground truth boundary. These FPs can be removed after the classification stage. In the literature, many researchers have devised automatic [18-21] as well as ensemble segmentation and classification [22-24] algorithms by combining several techniques to reduce the FPs at the detection stage. In general, the segmentation techniques can be characterized as thresholding based, region based (ie, region growing and region clustering), feature and edge based. We briefly summarize the advantages and disadvantages of segmentation techniques in Table 1.

**Figure 4.** Pixel-level illustration of true positive, false positive, and false negative compared with ground truth.

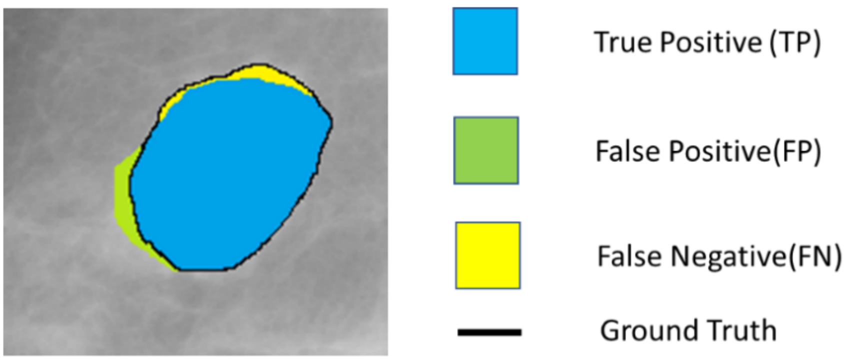


**Table 1.** Summary of advantages and disadvantages of segmentation methods.

| Methods | Advantages | Disadvantages |
|---|---|---|
| GT[a] | Widely used as preprocessing step in image processing as these methods are easy to implement | Not suitable for segmentation of ROIs[b], as GT methods produce high false positive detections |
| Local thresholding | Works well compared with GT, sometimes used to improve the GT results | Widely used in literature as initialization step of other algorithms, but local thresholding fails to separate the pixels accurately into suitable regions |
| Region growing | Uses pixel connectivity properties to grow iteratively and sum up the region having similar pixel properties | Need initialization point, that is, a seed point to begin with and highly dependent on initial guess |
| Region clustering | No seed point required to initialize; it can directly search the cluster regions. | Total number of clusters need to be predefined at initial stage |
| Edge detection | Highly suitable for detecting the object boundaries and contours of the suspected ROIs | Requires some information about object properties |
| Template matching | Needs ground truth and are easily implemented. Easy implementation; if the prototypes are suitably selected, it can produce good results. | Need prior information about the region properties of the objects such as size, shape, and area. |
| Multiscale technique | Do not require any prior knowledge about object properties | Requires empirical evaluation to select the appropriate wavelet transform |
| | Easily discriminate among the coefficients at different level and scale of decompositions | Need to select scale of decompositions |

[a]GT: Global thresholding.

[b]ROI: region of interest.

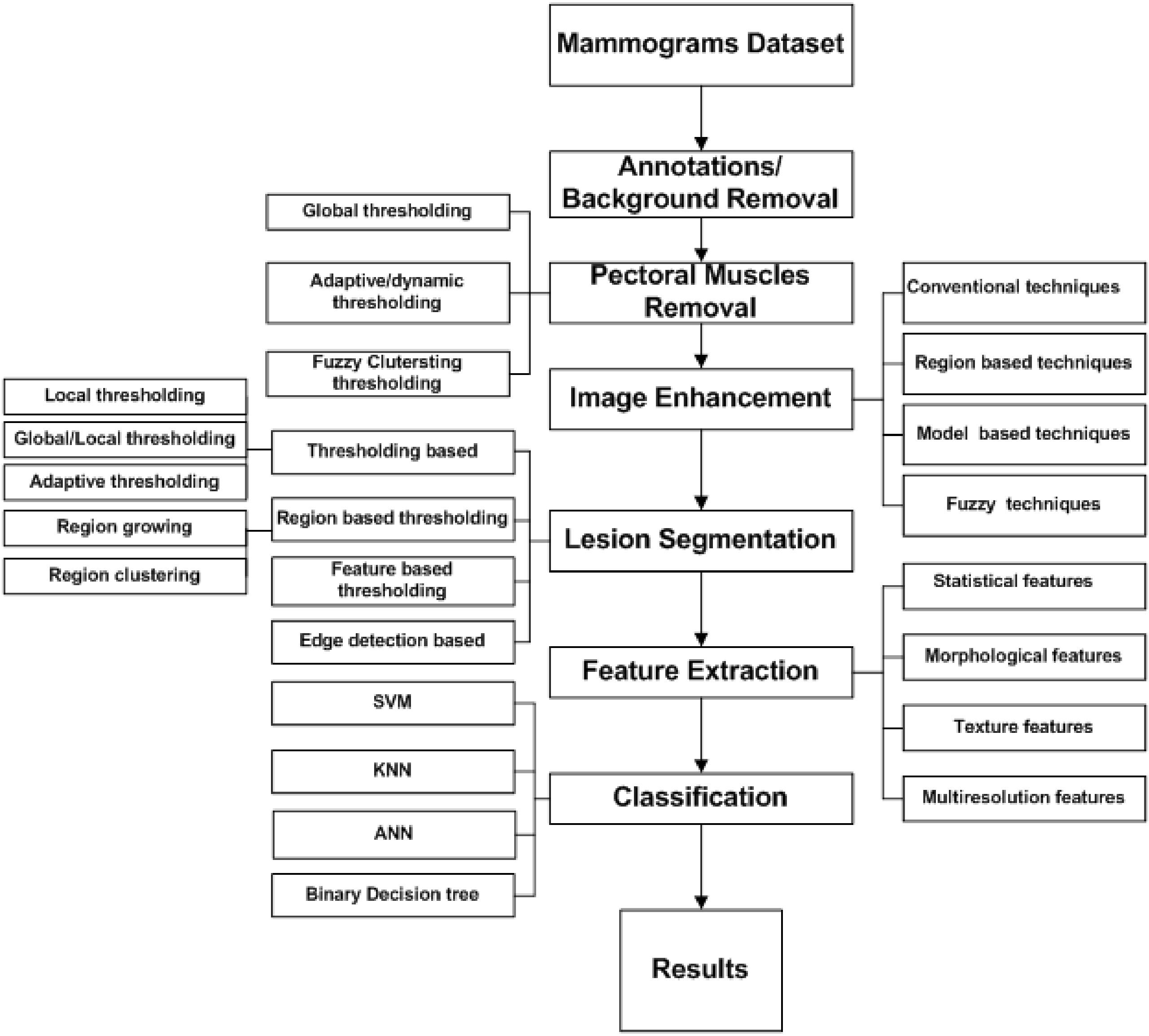
## Conventional Feature Extraction Techniques

In ML methods, learning the significant or most informative features from the medical images plays a vital role, as these features are used as discriminators in later stages for segmentation or classification. Most of these features are manually designed (handcrafted) based on clinicians' experience and prior knowledge about the target domain. Thus, the ML methods can be more problem oriented and often make it difficult for a nonexpert to exploit the full potential of the method. The feature extraction is the step that characterizes the features of a specific region. The significant features are retained for the classification step. To measure features from the ROIs, its properties such as mass size, regular or irregular shapes, homogeneity of boundaries, and density of tissues are utilized [25]. It is widely known that because of the variation in properties of normal and diseased tissues, feature space exhibits a large and complex nature. Most features are not significant when separately studied. However, when combined with other features, they can represent significant information that is helpful for the classification step. The performance of the algorithm is affected, and the complexity of the classifiers increases when excessive use of features is done. Thus, drawing the optimal features from images is very crucial. A number of feature selection techniques such as principal component analysis (PCA) [26], linear discriminant analysis (LDA) [27], filtering techniques such as chi-square test [28,29], and many other feature reduction methods [30] are used to select the most discriminative features to avoid overfitting and reduce the redundancy in feature space. On the basis of the feature characteristics, the feature space can be divided into 3 categories: morphological (shape or geometric) features, texture or statistical features, and multiresolution features.

## Classification Techniques

Classification is the last step to determine the lesion under observation is normal or cancerous regions. If it is classified as a cancerous region, further classification is done to determine the pathology of cancer, ie, benign or malignant. The classification step itself is heavily dependent on other intermediate steps, especially segmentation and feature extraction. In breast cancer classification, some of the commonly used classifiers include support vector machine (SVM) [31-34], artificial neural network (ANN) [10,13,21,35,36], k-nearest neighbor (KNN) [37,38], binary decision tree [39], and simple logistic classifier [34,40]. The performance of the classifier can be improved using some feature selection method to remove the redundant features and keep only the most discriminative features. An overview of CAD system based on ML algorithms for breast cancer diagnosis using mammographic data is illustrated in Figure 5.

**Figure 5.** An overview of mammogram processing using computer-aided diagnosis based on machine learning algorithms.




## Summary of Machine Learning Methods

A substantial amount of research on breast mass, microcalcification detection, and classification can be found in literature [22-24,31,34,41-45]. Masses are more challenging to detect compared with microcalcifications because the mass features may be concealed or alike to those of normal breast parenchyma. Thus, the detection of masses is still an open challenge in breast cancer detection. We also note that masses greatly vary in size [15,46], which makes it more challenging to detect. Another major limitation of the conventional ML studies is that mass analysis has not been done by defining some suitable scale for the range of masses. By defining the range of sizes, mass regions can be approached at the coarsest scale of description. However, a more confined approach is required to detect the boundaries of masses. Moreover, the variations in widths, lengths, and density spiculations that are associated with cancerous lesions and the varying scales require a more rigorous characterization and analysis. Apart from mass detection, architecture distortion and the detection of bilateral asymmetry are also important research topics in mammograms [34]. The new developments must cope and overcome with the challenges that existing algorithms exhibit by improving the performance. Furthermore, commercial CAD systems have achieved a reasonable degree of effectiveness to detect masses and calcifications. Future work on CAD systems for breast cancer diagnosis should focus on improving the performance. Feature extraction is one of the important steps in developing a CAD system. A broad variety of features for the characterization of breast cancer have been developed in the past years. Hence, more researches seem to be necessary to measure features robustness that can produce a high classification accuracy rate. Selecting the optimal feature subset for supervised learning problems requires an exhaustive search. The discriminative power of features used in CAD systems varies. Although some are highly significant for the discrimination of mammographic lesions, others are redundant or even irrelevant. Hence, automatic extraction of a subset of features from a higher dimensional feature vector is a common module in mammography CAD approaches.

## Deep Learning, an Overview

DL algorithms have made significant improvements in performance compared with other traditional ML and artificial intelligence [47]. The applications of DL have grown tremendously in various fields such as image classification [47], natural language processing [48], gaming [49]; and, in particular, it has become very popular in the medical imaging community for detection and diagnosis of diseases such as skin cancer [50,51], brain tumor detection, and segmentation [52].

The DL architectures can be characterized into 3 categories: unsupervised DL networks, also known as generative networks; supervised networks or discriminative networks; and hybrid or ensemble networks.

Convolutional neural network (CNN) is a state-of-the-art DL technique that is comprising many stacked convolutional layers [47]. The most common CNN discriminative architecture contains a convolutional layer, a maximum pooling layer to increase the field of view of the network, a rectified linear unit (ReLU), batch normalization, a softmax layer, and fully connected layers. The layers are aligned on top of each other to form a deep network that can the local and spatial information from this layer when a 2D or 3D image is presented as an input [53].

The AlexNet [47] architecture was one of the first deep networks for improving the ImageNet classification accuracy by a significant stride than the existing traditional methodologies. The architecture contained 5 convolutional layers proceeded by 3 fully connected layers. The ReLU activation function for the nonlinear part was introduced by replacing the traditional activation function such as Tanh or Sigmoid functions used in neural networks. ReLU has fast convergence as compared to sigmoid, which suffers from the vanishing gradient problem.

Later, VGG 16 architecture was proposed by visual geometry group (VGG) [54], Oxford University. The VGG improved the AlexNet architecture by changing the kernel size and introduction of multiple filters. The large kernel-sized filters are replaced (ie, 11×11 in Conv1 and 5×5 in Conv2, respectively) by multiple 3×3 kernel-sized filters that are placed one after another. The multiple smaller kernel filters improve the receptive field compared with a larger size kernel, as multiple nonlinear layers increase the depth of the network. The increased depth enables to learn more complex features at a lower cost. Although VGG achieved very good accuracy on classification tasks for the ImageNet dataset, it is computationally expensive and requires huge computational power, both in terms of storage memory and time. Thus, making it inefficient because of the large width of convolutional layers.

The GoogleNet [55] proposed the idea that most of the connection in dense architecture and their activations in the deep network are redundant or unnecessary due to correlations between them. This makes the network computationally expensive. Therefore, GoogleNet implied to have a most efficient network with sparse connections between the activations. GoogleNet introduced the inception module, which effectively computes sparse activation in a CNN with a normal dense construction. The network also uses 3 different convolutions sizes (ie, 5×5, 3×3, and 1×1) to have a better receptive field and extract details from very small levels. One of the important salient points about the inception module is that it also has a so-called bottleneck layer (1×1 conv.) that helps in massive reduction of the computation requirement. Another change that GoogleNet introduced is global average pooling at the last convolutional layer, thus averaging the channel values across the 2D feature map. This results in a reduction of the total number of parameters.

With increasing network depth, the accuracy of the network is saturated and thus degrades rapidly. This degradation is not caused by overfitting problem, but with the addition of more layers, the training error also increases that leads to degradation problem. The degradation problem was solved by introducing the residual network (ResNet) by He et al [56]. The residual module was introduced to effectively learn the training parameters in a deeper network. They introduced skip connections in convolutional layers in a blockwise manner to construct a residual module. The performance of ResNet is better than VGG and GoogleNet [57].

## Deep Learning for Breast Cancer Diagnosis

Many researchers have used DL approaches in medical image analysis. The success of DL is largely depending on the availability of large number of training samples to learn the descriptive feature mappings of the images, which give very accurate results in classification. For example, the image classification task, the network is trained over more than 1 million images with more than 1000 class data. However, in the case of medical images, the amount of available training data is not that big in size. Moreover, it is also difficult to acquire a large number of labeled images, as the annotation itself is an expensive task and for some diseases (eg, lesions) are scarce in the datasets [58]. In addition, annotation of these data samples, if exist, in different classes suffers from intraobserver variations, as the annotation is highly subjective and relies on the expert's knowledge and experience. To overcome the data insufficiency challenge, many research groups have devised different strategies: (1) using 2D patches or 3D cubes instead of using the whole image as input [59,60], which also reduces the model parameters and alleviates overfitting; (2) by introducing data augmentation using some affine transformations (translation, rotation, and flipping [61,62]) and training the network on the augmented data; (3) by transferring learning approach using pretrained weights [63,64] and just replacing the last layers by the new targeted class instead; and (4) using trained models with small input sizes and then transforming the weights in the fully connected layers into convolutional kernels [65].

## Search Strategy for Study Selection

To select the relevant recent studies on breast cancer diagnosis, we consider the studies in the past 5 years from well-known publishing platforms such as PubMed, Google Scholar, MEDLINE, Science Direct, Springer, and Web of Science databases. The search terms convolutional neural networks, deep learning, breast cancer, mass detection, transfer learning, and multiview are combined.

## Results

### Convolutional Neural Networks for Breast Cancer Diagnosis

In this section, we first present the methods that used breast density estimation methods as a tool for early diagnosis. Second, the methods used transfer learning and image features classification of suspected lesions into mass and normal class. Finally, we present the segmentation methods using semantic features for localization of masses and classifying the pathology.

### Convolutional Neural Network for Breast Density Estimation

Mammographic density is an important indicator of early breast cancer detection. In the United States, more than 30 states have agreed to use breast density as an earlier risk marker for cancer screening programmes [66]. The qualitative assessment is highly subjective, and there are wide variations in scoring results among the radiologists [67]. Recent studies also reveal that commercial software to assess the breast score tends to give mixed results by either over or under reporting when compared with assessment by radiologists [68,69]. The DL algorithms for density assessment can significantly reduce the burden of manual scoring for the radiologist and improve the performance for risk assessment [66].

One such attempt has been made by Mohamed et al [70] using CNN to classify the DMs based on breast densities. The Breast Imaging Reporting and Data System (BI-RADS) characterizes the densities into 4 classes. The discrimination of this breast densities acts as a risk marker for breast cancer, and radiologist can visually access the results. The study is focused on distinguishing the 2 difficult categories: scattered dense and heterogeneous dense breast tissues. Their method showed promising results for classification.

In another study, Ahn et al [71] presented a CNN-based approach for breast density estimation. The CNN was trained to learn image features from the image patches extracted from the whole mammograms and classify them as fatty and dense class tissues. The local and global statistical features were used to train the CNN. Wu et al [72] presented the application of deep neural network (DNN) for classification of breast densities in DMs. The study comprised 20,000 screening mammograms labeled as 4 class breast densities (ie, fatty, fibro-glandular dense, heterogeneously dense, and extremely dense). A scratch-based CNN with dense convolutional layers was used to discriminate the breast densities in the multiview data.

In a similar study, Xu et al [73] classified the breast density estimation method using residual CNN. The method worked efficiently for both single and multiview images (ie, CC and MLO). Their study aimed to use the residual CNN to discriminate the BI-RADS densities into 4 categories. The residual CNN consisted of 70 layers with 7 residual learning blocks. In addition, 2 other networks with 36 and 48 weighted layers but less residual blocks were also trained to compare the performance. The ResNets could minimize the cross-entropy loss to maximize classification accuracy. Their results showed that with increased residual layer, the classification accuracies improved. However, the computational cost was increased.

Kallenberg et al [74] proposed an unsupervised DL technique to classify the breast density and risk score in the segmented regions. The method uses conventional sparse autoencoder (CSAE) for learning the features. For mammographic density score, 3 class labels were used: PMs, fatty breast tissues, and dense breast tissues. For the mammographic texture score, 2 classes were considered (ie, cancer and normal patches). This score was used as a threshold to segment that tissue from the breast. Dice score showed the goodness of segmented versus the ground truth. The CSAE model was trained and tested for 3 different datasets, and the results showed a positive relationship with the scores obtained manually by experts.

Ionescu et al [75] proposed a CNN-based density estimation method to assist the radiologist in risk scoring. The CNN is trained to assess the visual analog score from unseen images. The method showed a strong correlation and match concordance indices results when compared with 2 independent readers in a clinical environment.

Geras et al [76] in their study used deep convolutional neural network for prediction of breast densities in multiview data. The method predicted breast density and classified into 3 types: BI-RADS0, BI-RADS1, and BI-RADS2. Moreover, it also classified the abnormalities from the ROIs extracted from these images into benign and malignant. The study also investigated the impact of training size and image size on prediction of accuracy. It was concluded that higher number of training samples improve the prediction accuracy during testing phase. Moreover, rescaling the image size did not have much effect on prediction accuracy of the method. The results show good agreement with manual scores done by expert radiologists.

Summary of the aforementioned methods is presented in Table 2 along with performance metrics for each method and the datasets used in these studies.

**Table 2.** Summary of convolutional neural network–based methods for breast density estimation.

| Author | Method | Dataset/number | Task | Performance metric/s (value/s) | Code availability |
|---|---|---|---|---|---|
| Mohamed et al [70] | CNN[a] (AlexNet; transfer learning) | Private, University of Pittsburgh/200,00 DM[b] (multiview) | Breast density estimation | AUC[c] (0.9882) | —[d] |
| Ahn et al [71] | CNN (transfer learning) | Private, Seoul University Hospital/397 DM (multiview) | Breast density estimation | Correlation coefficient (0.96) | — |
| Xu et al [73] | CNN (scratch based) | Public, INbreast dataset/410 DM (multiview) | Breast density estimation | Accuracy (92.63%) | — |
| Wu et al [72] | CNN (transfer learning) | Private, New York University School of Medicine/201,179 cases (multiview) | Breast density estimation | Mean AUC (0.934) | [77] |
| Kallenberg et al [74] | Conventional sparse autoencoder, ie, CNN+stacked autoencoder | Private, Dutch Breast Cancer Screening Program and Mayo Mammography, Minnesota/493+668 images (multiview) | Breast density estimation and risk scoring | Mammographic texture (0.91) and AUC (0.61) | — |
| Ionescu et al [75] | CNN | Private dataset/67,520 DM (multiview) | Breast density estimation and risk scoring | Average match concordance index of (0.6) | — |
| Geras et al [76] | Multiview deep neural network | Private, New York University/886,000 image (multiview) | Breast density estimation and risk score | Mean AUC (0.735) | — |

[a]CNN: convolutional neural network.

[b]DM: digital mammogram.

[c]AUC: area under the curve.

[d]Not available.

## Convolutional Neural Network for Breast Mass Detection

The automatic detection of masses at an early stage in DMs is still a hot topic of research. DL has significantly overcome the shortcomings of conventional methods by learning the object features. The learning curves of the DL methods have enabled to highlight the most relevant ROIs in DMs. In this section, we present the recent CNN-based methods for the detection of masses in mammograms using transfer learning techniques and scratch-based end-to-end training.

To improve the diagnostic accuracy of the breast CAD system, Dhungel et al [78] introduced a CAD system with minimal user intervention for breast mass detection, segmentation, and classification of the masses. The mass detection is done by cascade DL and random forest model for possible suspected regions that are refined by Bayesian optimization technique. The deep classifier is pretrained with regression analysis and handcrafted features, and the network in fine-tuned bases of ground truths for breast mass classification data, in particular, INbreast dataset, was used for experimentation. Although the method achieved significant results, one of the limitations of this method is that it requires fine-tuning at 2 stages. In addition, it was tested on limited images.

In another study, Dhungel et al [79] proposed a hybrid method for mass segmentation. The proposed conditional random field (CRF) model comprised several potential functions and a DL module for segmentation of masses in mammographic images. The method used tree reweighted (TRW) belief propagation method as a learning mechanism to reduce the lesion segmentation errors and provide optimal results. The study was performed on 2 multiview datasets (ie, INbreast and Digital Database for Screening Mammography [DDSM] datasets). Their results demonstrated that the DL module could improve the classification accuracies when combined with TRW.

Zhu et al [80] proposed a deep structural network with end-to-end learning for the segmentation of masses in DMs. The multistage deep network used a fully convolutional network (FCN) to model a potential function combined with a CRF to perform structured learning. FCN+CRF was used to obtain the empirical estimation of ROIs using the position prior information. To improve the predicted mass estimates, an adversarial training was introduced, which helped to eliminate the overfitting of mass regions with a smaller size in the mammogram dataset. The proposed multistage end-to-end network was evaluated on publicly available datasets (ie, INbreast and DDSM). The results demonstrate the effectiveness of that method.

In another study, Wang et al [81] presented a semiautomated early detection approach using DL to discriminate the microcalcifications and masses in breast cancer dataset. The method aimed to detect the microcalcifications that can be used as an indicator of early breast cancer [82,83]. The DL architecture consisted of stacked autoencoders (SAE) that stack multiple autoencoders, hierarchically. The deep SAE model used layer-wise greedy search training to extract the low-level semantic features of microcalcifications. The method had 2 scenarios: (1) having microcalcification and (2) microcalcifications and masses together to train and test the SAE model. Their method achieved good discriminative accuracy for identifying calcifications using SVM classifier.

Riddli et al [84] used transfer learning to implement the Faster R-CNN model for the detection of mammographic lesions and classify these lesions into benign and malignant pathology as can be seen in Figure 6 (adapted from [84]). The region proposal network in the Faster R-CNN generated possible suspected regions, which were refined by fine-tuning the hyperparameters. The method achieved significant classification results on the public INbreast database. However, one of the major limitations of this study is that it was tested on a small-scale pixel-level annotated data for detection, whereas the classification task was evaluated on a larger screening dataset.

Singh et al [85] presented a conditional generative adversarial network (cGAN) to segment mammographic masses from a ROI. The generative model learns the lesion representations to create binary masks. Although the adversarial network learns features that discriminate the real masses from the generated binary masks, the key advantage of their proposed cGAN is that it can work well for small sample dataset. The results of their method showed high similarity coefficient value and intersection over union of predicted masses with ground truths. Moreover, the method also classified the detected masses into 4 types (ie, irregular, lobular, oval, and round using CNN) as shown in Figure 7 (adapted from [85]).

Some researchers used image features for lesion detection and classification. One such study by Agarwal and Carson [86] predicted the semantic features such as the type of lesion and pathology in mammograms using the deep CNN. The motivation of the study was to propose a method that could automatically detect lesion and its pathology (ie, calcification or mass either benign or malignant). A scratch-based CNN was trained on DDSM dataset that contained mass as well as calcification cases. The method showed significant results in recognizing the semantic characteristics that can assist the radiologists in clinical decision support task.

Gao et al [87] presented a shallow-deep CNN (SD-CNN) for lesion detection and classification for contrast-enhanced DMs (CEDM). A 4-layered shallow-deep CNN was used to extract the visualization mappings of the convolutional layer in the CEDM images and combine them with low-energy (LE) images. This virtual enhancement improved the quality of LE images. ResNet was applied to these virtual combined images to extract the features to classify benign and normal cases. Using the SD-CNN on the CEDM images resulted in a significant improvement in classification accuracy compared with DMs.

Hagos et al [88] presented a multiview CNN to detect breast masses in symmetrical images. The method used CNN architecture with multiple patches as input to learn the symmetrical differences in the masses. Using the gradient orientation features and local lines on the images, the likelihood of pixels was used to determine the patch as mass or nonmass. They used the AUC and competition performance metric as performance measures for the proposed method against the baseline nonsymmetrical methods.

Later, Tuwen et al [89] proposed a multiview breast mass detection system based on DNN. The 2-step method first detected the suspicious regions in multiview data and then reduced FP through neural learning and affirmed the mass regions. The second major module consists of using transfer learning to train images with Fast R-CNN and mask R-CNN, with 3 different variants of ResNet (ie, ResNet-101, ResNeXt-101, and ResNeXt-152) as backend. The 3 networks were trained on full images to capture enough context information to discriminate soft lesion tissues. Data augmentation was also applied to enrich the dataset.

Jung et al [90] proposed a single-stage masses detection model using the RetinaNet model. RetinaNet is a 1-stage object detection method that can overcome the class imbalance problem and perform better than 2-stage methods. The focal loss function of the model allowed the RetinaNet to focus on the complex sample and detect objects. The mammogram RetinaNet was tested on 2 DM datasets, that is, INbreast and an in-house dataset GURO. Moreover, data augmentation was also used to enrich the database. Using the transfer learning approach, the mass patches from each image were trained using random weight initialization and a different combination. With 6 different experimental settings, the RetinaNet achieved significant detection accuracy compared with other state-of-the-art methods.

Shen et al [91] presented a deep architecture with end-to-end learning to detect and classify the mass regions in the whole digital breast image. The method was trained on the whole mammogram image by using a patch classifier to initiate weights of full image in an end-to-end fashion. The patch classifier uses existing VGG and ResNet architecture for classification. Different combinations of patch sets and hyperparameters were trained to find the optimal combination on whole breast images from the DDSM and INbreast datasets.

We summarize the lesion detection and classification methods in details in Table 3 and illustrate the datasets used, tasks, performance metrics, and code availability.

**Figure 6.** Sample results from the study by Ribli et al for mass detection and classification.

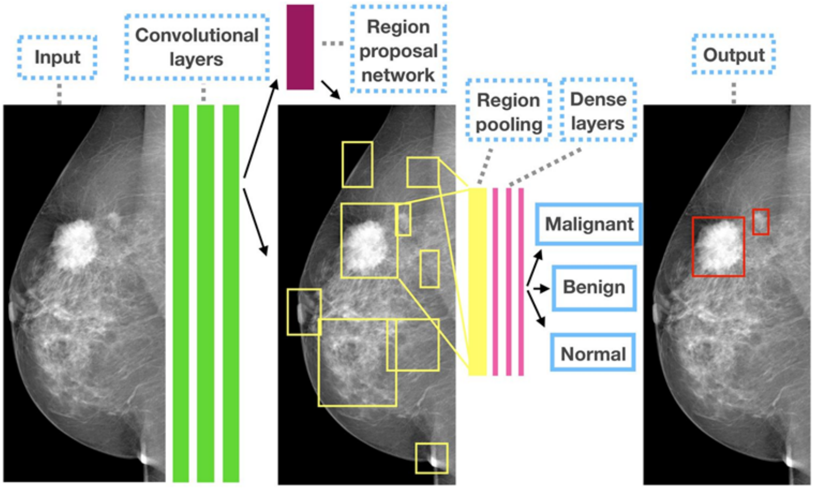



**Figure 7.** An overview of conditional generative adversarial network adapted from the study by Singh et al for mass segmentation and shape classification. CNN: convolutional neural network.

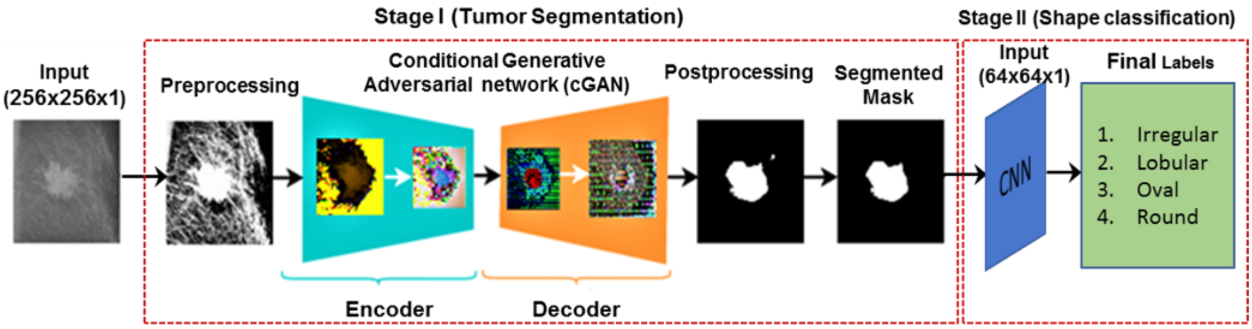


**Table 3.** Summary of convolutional neural network–based methods for breast mass detection.

| Author | Method | Dataset/number | Task | Performance metric/s (value/s) | Code availability |
|---|---|---|---|---|---|
| Dhungel et al [78] | Hybrid CNN[a]+level set | Public, INbreast dataset/410 images (multiview) | Mass detection, classification of benign, and malignant | Accuracy (0.9) and sensitivity (0.98) | —[b] |
| Dhungel et al [79] | CRF[c]+CNN | Public, INbreast and DDSM[d]/116 and 158 images (multiview) | Lesion detection and segmentation | Dice score (0.89) | — |
| Zhu et al [80] | Fully convolutional network+ CRF | Public, INbreast and DDSM/116 and 158 images (multiview) | Lesion segmentation | Dice score (0.97) | [92] |
| Wang et al [81] | Stacked autoencoder (transfer learning) | Private, Sun Yat-Sen University/1000 Digital mammogram | Detection and classification of calcifications and masses | Accuracy (0.87) | — |
| Riddli et al [84] | Faster R-CNN (transfer learning) | Public, DDSM (2620), INbreast (115), and private dataset by Semmelweis University Budapest/847 images | Detection and classification | AUC[e] (0.95) | Semmelweis dataset: [93]; Code: [94] |
| Singh et al [85] | Conditional generative adversarial network and CNN | Public and private, DDSM and Reus Hospital Spain dataset/567+194 images | Lesion segmentation and shape classification | Dice score (0.94) and Jaccard Index (0.89) | — |
| Agarwal and Carson [86] | CNN (scratch based) | Public, DDSM/8750 images (multiview) | Classification of mass and calcifications | Accuracy (0.90) | — |
| Gao et al [87] | Shallow-deep convolutional neural network, ie, 4 layers CNN+ResNet | Private, Mayo Clinic Arizona (49 subjects) and public, INbreast dataset (89 subjects) (multiview) | Lesion detection and classification | Accuracy (0.9) and AUC (0.92) | — |
| Hagos et al [88] | Multi-input CNN | Private (General Electric, Hologic, Siemens) dataset/28,294 images/(multiview) | Lesion detection and classification | AUC (0.93) and CPM (0.733) | — |
| Tuwen et al [89] | Fast R-CNN and Mask R-CNN with ResNet variants as backbone | Private (General Electric, Hologic, Siemens) dataset/23,405 images (multiview) | Lesion detection and classification | Sensitivity (0.97) with 3.56 FP[f] per image | — |
| Jung et al [90] | RetinaNet model | Public and private, INbreast and GURO dataset by Korea University Guro Hospital/410+222 images (multiview) | Mass detection and classification | Accuracy (0.98) with 1.3 FP per image | [95] |
| Shen et al [91] | CNN end-to-end (transfer learning through visual geometry group 16 and ResNet) | Public, DDSM and INbreast/2584 +410 (multiview) | Classification of masses | AUC (0.96) | [96] |

[a]CNN: convolutional neural network.

[b]Not available.

[c]CRF: conditional random field.

[d]DDSM: Digital Database for Screening Mammography.

[e]AUC: area under the curve.

[f]FP: false positive.

## Convolutional Neural Network features for Mass Classification

DL algorithms have shown significant improvements in breast cancer detection and classification problem over the past decade. The deep contextual and texture features allow the classifiers to discriminate between normal and abnormal lesions with varying shapes, size, and orientations. This not only improved the diagnostic capabilities of CAD system but also provided robust solutions for clinical practices.

Levy and Jain [97] demonstrated the usefulness of DL as a classification tool to discriminate the benign and malignant cancerous regions. The authors used a transfer learning approach to implement 2 architectures: AlexNet and GoogleNet. Data augmentation is used to increase the number of samples and alleviate overfitting issues. The results showed the significance of DL features in the classification of 2 classes.

Recently, Samala et al [98] presented mass classification method for digital breast tomosynthesis (DBT) using multistage fine-tuned CNN. The method used multistage transfer learning approach using different layer variation and selecting the optimal combination. Initially, the CNN tuned on ImageNet dataset was directly implemented on DBT data, and results were recorded in the multistage CNN that was fine-tuned on DBT dataset. The classification layers of CNN were used with different freeze pattern to extract the best combination that produces the highest accuracy. A total of 6 different combinations of transfer networks with varying freeze pattern for convolutional layers were tested. The multistage transfer learning significantly improved the results with least variations compared with single-stage learning.

Jadoon et al [99] presented a hybrid methodology for breast cancer classification by combining CNN with wavelet and curvelet transform. This model targeted a 3-class classification study (ie, normal, malignant, and benign cases). In this study, 2 methods, namely, CNN-discrete wavelet (CNN-DW) and CNN-curvelet transform (CNN-CT) were used. Features from wavelet and curvelet transform were fused with features obtained from the CNN. Data augmentation was used to enrich the dataset and avoid overfitting of features at the classification stage. Features from CNN-DW and CNN-CT were extracted at 4-level sub-band decompositions separately using the dense scale-invariant features at each sub-band level. The obtained features were presented as input to train a CNN with SoftMax and SVM layer for the classification of normal, benign, and malignant cases.

In a similar study, Huynh et al [100] also used transfer learning and CNN as tools to classify the tumors in breast cancer. The authors proposed an ensemble method that used both CNN and handcrafted features (eg, statistical and morphological features). The features from each method were combined to obtain the ensemble feature matrix. SVM classifier was used with 5-fold cross-validations. Performance of individual methods was compared with the ensemble method using 219 breast lesions. Their results showed that the ensemble could produce better results compared with fine-tuned CNN and analytical feature extractor.

Domingues and Cardoso [101] used an autoencoder to classify the mass versus not mass in the INbreast dataset. The classifier architecture included 1025-500-500-2000-2 layers with the same number of layers for the decoder as well. Except for the last 2 linear layers, all other layers were logistic. The method produced significant results. Moreover, it was also observed that increasing the depth of the network by adding more layers can also improve the detection and classification rates. The authors tested the performance of DL method against 5 classifiers (ie, KNN, decision trees, LDA, Naive Bayes, and SVM).

Wu et al [102] presented a DL approach to address the class imbalance and limited data issues for breast cancer classification. The approach used the infilling approach to generate synthetic mammogram patches using cGAN network. In the first step, the multiscale generator was trained to create synthetic patches in the target image using GAN. The generator used a cascading refinement to generate the multiscale features to ensure stability at high resolution. Figure 8 shows the synthetic images generated by cGAN. The cGAN was restricted to infill only lesion either mass or calcifications. The quality of generated images was experimentally evaluated by training a ResNet-50 classifier. The classification performance of cGAN augmented, and traditional augmentation methods were also compared. The results showed that synthetic augmentation improves classification.

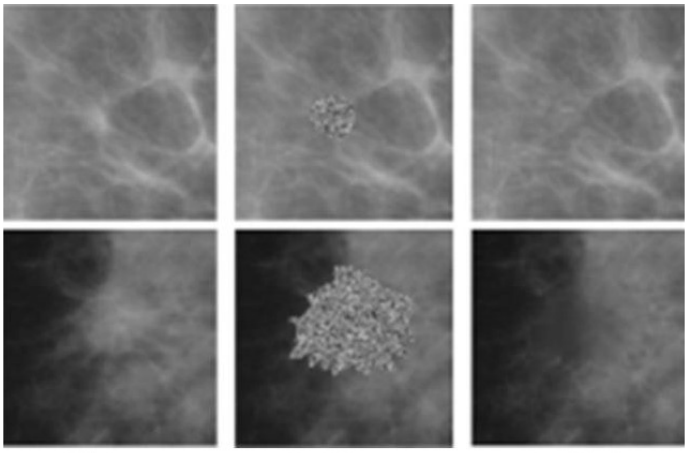
Sarah et al [103] addressed the issue of reducing the recall rates in breast cancer diagnosis. The higher number of FP results in higher recalls, which leads to unnecessary biopsies and increased cost for the patients. In this study, a DL method to reduce the recall rates was proposed. A deep CNN, namely, AlexNet, was implemented. A total of 6 different scenarios of mammogram classification were investigated. CNN was able to discriminate and classify these 6 categories very efficiently. Moreover, it could also be inferred that some features in recalled benign images classify them reexamined and to be recalled instead of classifying them as negative (normal) cases.

Lately, Wang et al [104] presented a hybrid DL method for multiview breast mass diagnosis. The framework exploited the contextual information from the multiview data (ie, CC and MLO) using CNN features and attention mechanism. The proposed multiview DNN aimed to help medical experts for the classification of breast cancer lesion. The method comprised 4 steps, and mass cropping and extraction of clinical features were done from the multiview patches. The recurrent neural network, in particular, long short-term memory, was used to extract the label co-occurrence dependency of multiview information for the classification of mass regions into benign and malignant cases using the clinical and CNN features as input.

In another study, Shams et al [105] proposed a GAN-based mammogram classification method—Deep GeneRAtive Multitask (DiaGRAM) network to deal with data scarcity and limited availability of annotated data. The DiaGRAM effectively uses an end-to-end multitask learning to improve diagnostic performance on limited number of datasets.

Gastitouni et al [106] presented an ensemble method for breast pectoral parenchymal classification. The texture feature maps extracted from lattice-based techniques are fed as input separately to a multichannel CNN. The meta-features from the CNN predicted the risk score associated with breast parenchyma. The hybrid method showed better performance compared with individual texture features and CNN, respectively.

**Figure 8.** Sample results from the study by Wu et al for synthetic generation of data using conditional generative adversarial network. GAN: generative adversarial network.




Dhungel et al [107] introduced a multiview ensemble deep ResNet (mResNet) for classification of malignant and benign tumors. Their ensemble network comprised deep ResNet capable to tackle 6 input images, with different views, that is, CC and MLO. The mResNet can automatically produce binary maps of the lesions. The final output of the mResNet are concatenated to obtain a fully connected layer that can classify the lesions into malignant or benign class.

Generally, DL methods have significantly improved the performance of breast cancer detection, classification, and segmentation. We summarize these methods in details in Table 4 and illustrate the datasets used, tasks, performance metrics, and code availability.

**Table 4.** Summary of convolutional neural network–based methods for breast mass classification.

| Author | Method | Dataset/number | Task | Performance metric/s (value/s) | Code availability |
|---|---|---|---|---|---|
| Levy and Jain [97] | AlexNet and GoogleNet (transfer learning) | Public, DDSM[a] dataset/1820 images (multiview) | Breast mass classification | Accuracy (0.924), precision (0.924), and recall (0.934) | —[b] |
| Samala et al [98] | Multistage fine-tuned CNN[c] (transfer learning) | Private+public, University of Michigan and DDSM/4039 ROIs[d] (multiview) | Classification performance on varying sample sizes | AUC[e] (0.91) | [108] |
| Jadoon et al [99] | CNN- Discrete wavelet and CNN-curvelet transform | Public, image retrieval in medical applications dataset/2796 ROI patches | Classification | Accuracy (81.83 and 83.74) and receiver operating characteristic curve (0.831 and 0.836) for both methods | — |
| Huynh et al [100] | CNN (transfer learning) | Private, University of Chicago/219 images (multiview) | Classification of benign and malignant tumor | AUC (0.86) | — |
| Domingues and Cardoso [101] | Autoencoder | Public, INbreast/116 ROIs | Classification of mass vs normal | Accuracy (0.99) | [109] |
| Wu et al [102] | GAN[f] and ResNet50 | Public, DDSM dataset/10,480 images (multiview) | Detection and classification of benign and malignant calcifications and masses | AUC (0.896) | [110] |
| Sarah et al [103] | CNN (transfer learning) | Public, Full-field digital mammography and DDSM/14,860 images (multiview) | Classification | AUC (0.91) | — |
| Wang et al [104] | CNN and long short-term memory | Public, Breast Cancer Digital Repository (BCDR-F03)/763 images (multiview) | Classification of breast masses using contextual information | AUC (0.89) | — |
| Shams et al [105] | CNN and GAN | Public, INbreast and DDSM (multiview) | Classification | AUC (0.925) | — |
| Gastounioti et al [106] | Texture feature+CNN | Private/106 cases (mediolateral oblique view only) | Classification | AUC (0.9) | — |
| Dhungel et al [107] | Multi-ResNet | Public, INbreast (multiview) | Classification | AUC (0.8) | — |

[a]DDSM: Digital Database for Screening Mammography.

[b]Not available.

[c]CNN: convolutional neural network.

[d]ROIs: region of interest.

[e]AUC: area under the curve.

[f]GAN: generative adversarial network.

## Discussion

### Principal Findings

From Tables 2, 3, and 4, it can be noted that significant works have been done on breast cancer diagnosis. The review of breast diagnosis methods shows that DL has helped to improve the diagnostic performance of the breast CAD system, but still challenges remain for clinical applicability of such methods, and more research is needed. The presented literature aims to help in building a CAD system that is robust, computationally efficient to assist the clinicians in the diagnosis of breast cancer at early stages. One main problem related to mammograms is the heterogeneity of breast tissues; that is, the images acquired at CC and MLO view may not show with different densities. Some researchers use breast density estimation scores as the initial biomarker for the presence of cancer. However, the analysis shows that these methods can be confined to a particular type of breast density and cannot be generalized for the whole population. Others use DL in a hybrid approach and a semi supervised manner to extract significant semantic and contextual information to detect and classify the breast lesions.

On the other hand, many attempts have been made to reduce human intervention and produce fully automatic CAD system, which is a very challenging task. In fact, all methods in literature require annotated images (ground truth) to validate their findings during the training and testing stages. Thus, acquisition of labeled mammograms with image-level and pixel-level annotations is one of the obstacles in designing robust DL methods. The main issue is not only the availability of data but also annotations by expert radiologist, which is time consuming, subjective, and expensive.

It is noted from the literature that the automated DL method requires extensive experimentation, computational power, and preprocessing of data, which make it inefficient to be used in real time. Moreover, finding the optimal parameters in DL networks is also one of the major challenges in building a CAD system for clinical use. However, this issue can be resolved if sufficient training is provided to clinicians, and CAD systems are made more user friendly. It is also noted that the semisupervised approaches have shown good performance on the public and private datasets for breast cancer diagnosis.

From the analysis of methods mentioned in Tables 2, 3, and 4, it can be noted that most methods mentioned previously adapt the augmentation strategies to enrich the dataset. All these techniques only use geometric transformations to create rotated and scale version of existing samples without adding any morphological variations in the lesions. Thus, enrichment of data with more samples is only limited to affine transformations and cannot fully resolve the overfitting problem.

Developing DL models that can learn from limited data is still an open research area not only in breast cancer diagnosis but also for other medical image analysis applications. Moreover, developing data augmentation techniques that can create morphological variations in augmented samples, while also preserving the lesion characteristic, are needed. One of the solutions to address these problems is to explore the capabilities of GANs as successfully demonstrated in studies by Singh et al [85] and Wu et al [102]. Techniques such as these will not only tackle the insufficiency issue of data but will also provide a viable solution to class imbalance problem, which is also an important research area.

Apart from the development of automatic DL techniques, there are other associated challenges to the medical imaging research community. First, it is very challenging to secure funding for construction of a medical dataset. Also, finding an expert for annotation and the cost of annotation itself is very high. Second, privacy and copyright issues make the medical image difficult to share compared with natural images datasets. Finally, because of the complex anatomy of human organs, a variety of dataset is required using different imaging modalities. Despite these challenges, there has been a significant increase in the number of public datasets. Organizing a grand challenge is one of the good practices devised to share and enrich the datasets. The participants are provided with a certain number of tasks on a particular dataset, and the technique with best results is announced as a winner. Moreover, different research centers join hands in research collaborations as well as common data sharing platforms.

## Conclusions

From the aforementioned discussions, we can see that both supervised and unsupervised DL methods are used by the image analysis community, but the majority of the work uses the semi supervised approach. The presented literature aims to help in building a CAD system that is robust and computationally efficient to assist the clinicians in the diagnosis of breast cancer at early stages. As DL requires a sufficient amount of annotated data for training, most of the researchers use a combination of public and private data followed by data augmentation techniques to overcome the data scarcity issue. These approaches have provided a feasible solution to the problem of scarcity of data and overfitting.

## Acknowledgments

This work was supported partly by National Natural Science Foundation of China (Nos 61871274, 61801305, 81771922, and 81571758), National Natural Science Foundation of Guangdong Province (Nos 2017A030313377 and 2016A030313047), Shenzhen Peacock Plan (Nos KQTD2016053112051497 and KQTD2015033016104926), and Shenzhen Key Basic Research Project (Nos JCYJ20170413152804728, JCYJ20180507184647636, JCYJ20170818142347251 and JCYJ20170818094109846).

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Commonly used metrics for performance evaluation in breast cancer diagnosis.

[PDF File (Adobe PDF File), 69KB - jmir_v21i7e14464_app1.pdf ]

## References

1. American Cancer Society. 2018. Global Cancer: Facts & Figures, 4th edition URL: http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/global-cancer-facts-and-figures/global-cancer-facts-and-figures-4th-edition.pdf

2. Blakely T, Shaw C, Atkinson J, Cunningham R, Sarfati D. Social inequalities or inequities in cancer incidence? Repeated census-cancer cohort studies, New Zealand 1981-1986 to 2001-2004. Cancer Causes Control 2011 Sep;22(9):1307-1318. [doi: 10.1007/s10552-011-9804-x] [Medline: 21717195]
3. Smigal C, Jemal A, Ward E, Cokkinides V, Smith R, Howe HL, et al. Trends in breast cancer by race and ethnicity: update 2006. CA Cancer J Clin 2006;56(3):168-183 [FREE Full text] [doi: 10.3322/canjclin.56.3.168] [Medline: 16737949]
4. Guide to Mammography And Other Breast Imaging Procedures. New York: Natl Council on Radiation; 2012.
5. Ponraj DN, Jenifer ME, Poongodi DP, Manoharan JS. A survey on the preprocessing techniques of mammogram for the detection of breast cancer. J Emerg Trends Comput Inf Sci 2011;2(12):656-664 [FREE Full text]
6. Rangayyan RM, Ayres FJ, Leo Desautels JE. A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs. J Franklin Inst 2007 May;344(3-4):312-348. [doi: 10.1016/j.jfranklin.2006.09.003]
7. Ganesan K, Acharya UR, Chua KC, Min LC, Abraham KT. Pectoral muscle segmentation: a review. Comput Methods Programs Biomed 2013 Apr;110(1):48-57. [doi: 10.1016/j.cmpb.2012.10.020] [Medline: 23270962]
8. Ge M, Mainprize JG, Mawdsley GE, Yaffe MJ. Segmenting pectoralis muscle on digital mammograms by a Markov random field-maximum a posteriori model. J Med Imaging (Bellingham) 2014 Oct;1(3):34503 [FREE Full text] [doi: 10.1117/1.JMI.1.3.034503] [Medline: 26158068]
9. Ali MA, Czene K, Eriksson L, Hall P, Humphreys K. Breast tissue organisation and its association with breast cancer risk. Breast Cancer Res 2017 Sep 6;19(1):103 [FREE Full text] [doi: 10.1186/s13058-017-0894-6] [Medline: 28877713]
10. Oliver A, Lladó X, Torrent A, Martí J. One-Shot Segmentation of Breast, Pectoral Muscle, and Background in Digitised Mammograms. In: Proceedings of the International Conference on Image Processing.: IEEE; 2014 Presented at: IEEE'14; October 27-30, 2014; Paris, France p. 912-916. [doi: 10.1109/ICIP.2014.7025183]
11. Bora VB, Kothari AG, Keskar AG. Robust automatic pectoral muscle segmentation from mammograms using texture gradient and euclidean distance regression. J Digit Imaging 2016 Feb;29(1):115-125 [FREE Full text] [doi: 10.1007/s10278-015-9813-5] [Medline: 26259521]
12. Tourassi GD, Vargas-Voracek R, Catarious Jr DM, Floyd Jr CE. Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. Med Phys 2003 Aug;30(8):2123-2130. [doi: 10.1118/1.1589494] [Medline: 12945977]
13. Rampun A, Morrow PJ, Scotney BW, Winder J. Fully automated breast boundary and pectoral muscle segmentation in mammograms. Artif Intell Med 2017 Dec;79:28-41. [doi: 10.1016/j.artmed.2017.06.001] [Medline: 28606722]
14. Eltoukhy MM, Faye I. An Adaptive Threshold Method for Mass Detection in Mammographic Images. In: Proceedings of the International Conference on Signal and Image Processing Applications. 2013 Presented at: IEEE'13; October 8-10, 2013; Melaka, Malaysia p. 374-378.
15. Gardezi SJ, Faye I, Sanchez BJ, Kamel N, Hussain M. Mammogram classification using dynamic time warping. Multimed Tools Appl 2017 Jan 10;77(3):3941-3962. [doi: 10.1007/s11042-016-4328-8]
16. Shih FY. Image Processing and Pattern Recognition: Fundamentals and Techniques. USA: Wiley-IEEE Press; 2010.
17. Biltawi M, Al-Najdawi N, Tedmori S. Mammogram Enhancement and Segmentation Methods: Classification, Analysis, and Evaluation. In: Proceedings of the 13th International Arab Conference on Information Technology. 2012 Presented at: ACIT'12; December 10-13, 2012; Jordan, Zarq p. 477-485.
18. de Oliveira HC, Mencattini A, Casti P, Martinelli E, di Natale C, Catani JH, et al. Reduction of False-Positives in a CAD Scheme for Automated Detection of Architectural Distortion in Digital Mammography. In: Proceeding of the Conference on Computer-Aided Diagnosis. 2018 Presented at: SPIE'18; February 10-15, 2018; Houston, Texas, United States p. 105752P. [doi: 10.1117/12.2293388]
19. Liu X, Zeng Z. A new automatic mass detection method for breast cancer with false positive reduction. Neurocomputing 2015 Mar;152:388-402 [FREE Full text] [doi: 10.1016/j.neucom.2014.10.040]
20. Jen CC, Yu SS. Automatic detection of abnormal mammograms in mammographic images. Expert Syst Appl 2015 Apr;42(6):3048-3055. [doi: 10.1016/j.eswa.2014.11.061]
21. Ayer T, Chen Q, Burnside ES. Artificial neural networks in mammography interpretation and diagnostic decision making. Comput Math Methods Med 2013;2013:832509 [FREE Full text] [doi: 10.1155/2013/832509] [Medline: 23781276]
22. Magna G, Casti P, Jayaraman SV, Salmeri M, Mencattini A, Martinelli E, et al. Identification of mammography anomalies for breast cancer detection by an ensemble of classification models based on artificial immune system. Knowl Based Syst 2016 Jun;101:60-70. [doi: 10.1016/j.knosys.2016.02.019]
23. Wang H, Zheng B, Yoon SW, Ko HS. A support vector machine-based ensemble algorithm for breast cancer diagnosis. Eur J Oper Res 2018 Jun;267(2):687-699 [FREE Full text] [doi: 10.1016/j.ejor.2017.12.001]
24. Sert E, Ertekin S, Halici U. Ensemble of Convolutional Neural Networks for Classification of Breast Microcalcification From Mammograms. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2017 Presented at: IEEE'17; July 11-15, 2017; Seogwipo, South Korea p. 689-692. [doi: 10.1109/EMBC.2017.8036918]
25. Cheng H, Shi X, Min R, Hu L, Cai X, Du H. Approaches for automated detection and classification of masses in mammograms. Pattern Recognit 2006 Apr;39(4):646-668. [doi: 10.1016/j.patcog.2005.07.006]

26. Hasan H, Tahir NM. Feature Selection of Breast Cancer Based on Principal Component Analysis. In: Proceedings of the 6th International Colloquium on Signal Processing & Its Applications. 2010 Presented at: IEEE'10; May 21-23, 2010; Mallaca City, Malaysia p. 1-4. [doi: 10.1109/CSPA.2010.5545298]
27. Chan HP, Wei D, Helvie MA, Sahiner B, Adler DD, Goodsitt MM, et al. Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space. Phys Med Biol 1995 May;40(5):857-876. [Medline: 7652012]
28. Jin X, Xu A, Bie R, Guo P. Machine Learning Techniques and Chi-square Feature Selection for Cancer Classification Using SAGE Gene Expression Profiles. In: Proceedings of the 2006 International Conference on Data Mining for Biomedical Applications. 2006 Presented at: BioDM'06; April 9, 2006; Singapore p. 106-115. [doi: 10.1007/11691730_11]
29. Salama GI, Abdelhalim MB, Zeid MA. Breast cancer diagnosis on three different datasets using multi-classifiers. Int J Comput Sci Inf Technol 2012;1(1):36-43 [FREE Full text]
30. Saghapour E, Kermani S, Sehhati M. A novel feature ranking method for prediction of cancer stages using proteomics data. PLoS One 2017;12(9):e0184203 [FREE Full text] [doi: 10.1371/journal.pone.0184203] [Medline: 28934234]
31. Eltoukhy MM, Gardezi SJ, Faye I. A Method to Reduce Curvelet Coefficients for Mammogram Classification. In: Proceedings of the Region 10 Symposium. 2014 Presented at: IEEE'14; April 14-16, 2014; Kaula lumpur , Malaysia p. 663-666. [doi: 10.1109/TENCONSpring.2014.6863116]
32. Singh B, Jain V, Singh S. Mammogram mass classification using support vector machine with texture, shape features and hierarchical centroid method. J Med Imaging & Health Infor 2014 Oct 1;4(5):687-696 [FREE Full text] [doi: 10.1166/jmihi.2014.1312]
33. Sonar P, Bhosle U, Choudhury C. Mammography Classification Using Modified Hybrid SVM-KNN. In: Proceedings of the International Conference on Signal Processing and Communication. 2017 Presented at: IEEE'17; July 28-29, 2017; Coimbatore, India p. 305-311. [doi: 10.1109/CSPC.2017.8305858]
34. Gardezi SJ, Faye I, Eltoukhy MM. Analysis of Mammogram Images Based on Texture Features of Curvelet Sub-Bands. In: Proceedings of the 5th International Conference on Graphic and Image Processing. 2014 Presented at: SPIE'14; January 10, 2014; Hong Kong p. 906924. [doi: 10.1117/12.2054183]
35. Pratiwi M, Alexander, Harefa J, Nanda S. Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network. Procedia Comput Sci 2015;59:83-91 [FREE Full text] [doi: 10.1016/j.procs.2015.07.340]
36. Pal NR, Bhowmick B, Patel SK, Pal S, Das J. A multi-stage neural network aided system for detection of microcalcifications in digitized mammograms. Neurocomputing 2008 Aug;71(13-15):2625-2634 [FREE Full text] [doi: 10.1016/j.neucom.2007.06.015]
37. Wu X, Kumar V, Ross QJ, Ghosh JR, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowl Inf Syst 2007 Dec 4;14(1):1-37 [FREE Full text] [doi: 10.1007/s10115-007-0114-2]
38. Tan PN, Steinbach M, Kumar V. Introduction to Data Mining. Boston, USA: Pearson Addison Wesley; 2006.
39. Sumbaly R, Vishnusri N, Jeyalatha S. Diagnosis of breast cancer using decision tree data mining technique. Int J Comput Appl 2014 Jul 18;98(10):16-24. [doi: 10.5120/17219-7456]
40. Landwehr N, Hall M, Frank E. Logistic model trees. Mach Learn 2005 May;59(1-2):161-205 [FREE Full text] [doi: 10.1007/s10994-005-0466-3]
41. Ramirez-Villegas JF, Ramirez-Moreno DF. Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions. Neurocomputing 2012 Feb;77(1):82-100. [doi: 10.1016/j.neucom.2011.08.015]
42. Wajid SK, Hussain A. Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. Expert Syst Appl 2015 Nov;42(20):6990-6999 [FREE Full text] [doi: 10.1016/j.eswa.2015.04.057]
43. Zhang X, Homma N, Goto S, Kawasumi Y, Ishibashi T, Abe M, et al. A hybrid image filtering method for computer-aided detection of microcalcification clusters in mammograms. J Med Eng 2013;2013:615254 [FREE Full text] [doi: 10.1155/2013/615254] [Medline: 27006921]
44. Abbas Q, Celebi ME, Garcı´a I. Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system. Biomed Signal Process Control 2013 Mar;8(2):204-214. [doi: 10.1016/j.bspc.2012.08.003]
45. Jiang M, Zhang S, Li H, Metaxas DN. Computer-aided diagnosis of mammographic masses using scalable image retrieval. IEEE Trans Biomed Eng 2015 Feb;62(2):783-792. [doi: 10.1109/TBME.2014.2365494] [Medline: 25361497]
46. Michaelson J, Satija S, Moore R, Weber G, Halpern E, Garland A, et al. Estimates of the sizes at which breast cancers become detectable on mammographic and clinical grounds. J Womens Health 2003;5(1):3-10 [FREE Full text] [doi: 10.1097/00130747-200302000-00002]
47. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 May 24;60(6):84-90. [doi: 10.1145/3065386]
48. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. IEEE Comput Intell Mag 2018 Aug;13(3):55-75. [doi: 10.1109/MCI.2018.2840738]
49. Schuurmans D, Zinkevich MA. Deep Learning Games. In: Proceedings of the Advances in Neural Information Processing Systems. 2016 Presented at: NPIS'16; December 5-10, 2016; Barcelona, Spain.

50. Yang X, Zeng Z, Yeo SY, Tan C, Tey HL, Su Y. Cornell University. 2017. A Novel Multi-task Deep Learning Model for Skin Lesion Segmentation and Classification URL: https://arxiv.org/abs/1703.01025
51. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 2;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
52. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. Med Image Anal 2017 Dec;35:18-31. [doi: 10.1016/j.media.2016.05.004] [Medline: 27310171]
53. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based Learning Applied to Document Recognition. In: Proceedings of the IEEE. 1998 Presented at: IEEE'98; May 20-22, 1998; Pasadena, USA p. 2278-2324. [doi: 10.1109/5.726791]
54. Simonyan K, Zisserman A. Cornell University. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition URL: https://arxiv.org/abs/1409.1556
55. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D. Going Deeper With Convolutions. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. 2015 Presented at: IEEE'15; June 7-12, 2015; Boston, MA, USA p. 1-9. [doi: 10.1109/CVPR.2015.7298594]
56. He K, Zhang X, Ren S, Sun J. Identity Mappings in Deep Residual Networks. In: Proceedings of the Conference on Computer Vision. 2016 Presented at: ECCV'16; October 11-14, 2016; Amsterdam, The Netherlands p. 630-645. [doi: 10.1007/978-3-319-46493-0_38]
57. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. 2016 Presented at: IEEE'16; June 27-30, 2016; Las Vegas, NV, USA p. 770-778. [doi: 10.1109/CVPR.2016.90]
58. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 2016 Dec;35(5):1299-1312. [doi: 10.1109/TMI.2016.2535302] [Medline: 26978662]
59. Suk HI, Lee SW, Shen D, Alzheimer's Disease Neuroimaging Initiative. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage 2014 Nov 1;101:569-582 [FREE Full text] [doi: 10.1016/j.neuroimage.2014.06.077] [Medline: 25042445]
60. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Rep 2016 Apr 15;6:24454 [FREE Full text] [doi: 10.1038/srep24454] [Medline: 27079888]
61. Ting FF, Tan YJ, Sim KS. Convolutional neural network improvement for breast cancer classification. Expert Syst Appl 2019 Apr;120:103-115 [FREE Full text] [doi: 10.1016/j.eswa.2018.11.008]
62. Setio AA, Ciompi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging 2016 Dec;35(5):1160-1169. [doi: 10.1109/TMI.2016.2536809] [Medline: 26955024]
63. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016 Dec;35(5):1285-1298 [FREE Full text] [doi: 10.1109/TMI.2016.2528162] [Medline: 26886976]
64. Shie CK, Chuang CH, Chou CN, Wu MH, Chang EY. Transfer Representation Learning for Medical Image Analysis. In: Proceedings of the Engineering in Medicine and Biology Society. 2015 Presented at: IEEE'15; August 25-29, 2015; Milan, Italy p. 711-714. [doi: 10.1109/EMBC.2015.7318461]
65. Qi D, Hao C, Lequan Y, Lei Z, Jing Q, Defeng W, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans Med Imaging 2016 Dec;35(5):1182-1195. [doi: 10.1109/TMI.2016.2528129] [Medline: 26886975]
66. Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, et al. Mammographic breast density assessment using deep learning: clinical implementation. Radiology 2019 Jan;290(1):52-58. [doi: 10.1148/radiol.2018180694] [Medline: 30325282]
67. Sprague BL, Conant EF, Onega T, Garcia MP, Beaber EF, Herschorn SD, PROSPR Consortium. Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study. Ann Intern Med 2016 Oct 4;165(7):457-464 [FREE Full text] [doi: 10.7326/M15-2934] [Medline: 27428568]
68. Youk JH, Gweon HM, Son EJ, Kim JA. Automated volumetric breast density measurements in the era of the BI-RADS fifth edition: a comparison with visual assessment. AJR Am J Roentgenol 2016 May;206(5):1056-1062. [doi: 10.2214/AJR.15.15472] [Medline: 26934689]
69. Brandt KR, Scott CG, Ma L, Mahmoudzadeh AP, Jensen MR, Whaley DH, et al. Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. Radiology 2016 Jun;279(3):710-719 [FREE Full text] [doi: 10.1148/radiol.2015151261] [Medline: 26694052]
70. Mohamed AA, Berg WA, Peng H, Luo Y, Jankowitz RC, Wu S. A deep learning method for classifying mammographic breast density categories. Med Phys 2018 Jan;45(1):314-321 [FREE Full text] [doi: 10.1002/mp.12683] [Medline: 29159811]
71. Ahn CK, Heo C, Jin H, Kim JH. A Novel Deep Learning-Based Approach to High Accuracy Breast Density Estimation in Digital Mammography. In: Proceedings of the Computer-Aided Diagnosis. 2017 Presented at: SPIE'17; February 11-16,

2017; Orlando, Florida, US URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10134.toc [doi: 10.1117/12.2254264]
72. Wu N, Geras KJ, Shen Y, Su J, Kim SG, Kim E. Breast Density Classification With Deep Convolutional Neural Networks. In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing. 2018 Presented at: IEEE'18; April 15-20, 2018; Calgary, Canada p. 6682-6686. [doi: 10.1109/ICASSP.2018.8462671]
73. Xu J, Li C, Zhou Y, Mou L, Zheng H, Wang S. Cornell University. 2018. Classifying Mammographic Breast Density by Residual Learning URL: https://arxiv.org/abs/1809.10241
74. Kallenberg M, Petersen K, Nielsen M, Ng AY, Pengfei D, Igel C, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans Med Imaging 2016 Dec;35(5):1322-1331. [doi: 10.1109/TMI.2016.2532122] [Medline: 26915120]
75. Ionescu GV, Fergie M, Berks M, Harkness EF, Hulleman J, Brentnall AR, et al. Prediction of reader estimates of mammographic density using convolutional neural networks. J Med Imaging (Bellingham) 2019 Jul;6(3):031405. [doi: 10.1117/1.JMI.6.3.031405] [Medline: 30746393]
76. Geras KJ, Wolfson S, Shen Y, Kim S, Moy L, Cho K. Cornell University. 2017. High-Resolution Breast Cancer Screening With Multi-View Deep Convolutional Neural Networks URL: https://arxiv.org/abs/1703.07047
77. GitHub. Breast density classification with deep convolutional neural networks URL: https://github.com/nyukat/breast_density_classifier
78. Dhungel N, Carneiro G, Bradley A. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med Image Anal 2017 Dec;37:114-128. [doi: 10.1016/j.media.2017.01.009] [Medline: 28171807]
79. Dhungel N, Carneiro G, Bradley AP. Tree RE-Weighted Belief Propagation Using Deep Learning Potentials for Mass Segmentation From Mammograms. In: Proceedings of the 12th International Symposium on Biomedical Imaging. 2015 Presented at: IEEE'15; April 16-19, 2015; New York, USA p. 760-763. [doi: 10.1109/ISBI.2015.7163983]
80. Zhu W, Xiang X, Tran TD, Hager GD, Xie X. Adversarial Deep Structured Nets for Mass Segmentation From Mammograms. In: Proceedings of the 15th International Symposium on Biomedical Imaging. 2018 Presented at: IEEE'18; April 4-7, 2018; Washington, DC p. 847-850. [doi: 10.1109/ISBI.2018.8363704]
81. Wang J, Yang X, Cai H, Tan W, Jin C, Li L. Discrimination of breast cancer with microcalcifications on mammography by deep learning. Sci Rep 2016 Dec 7;6:27327 [FREE Full text] [doi: 10.1038/srep27327] [Medline: 27273294]
82. Fletcher SW, Elmore JG. Clinical practice. Mammographic screening for breast cancer. N Engl J Med 2003 Apr 24;348(17):1672-1680 [FREE Full text] [doi: 10.1056/NEJMcp021804] [Medline: 12711743]
83. Yunus M, Ahmed N, Masroor I, Yaqoob J. Mammographic criteria for determining the diagnostic value of microcalcifications in the detection of early breast cancer. J Pak Med Assoc 2004 Jan;54(1):24-29 [FREE Full text] [Medline: 15058638]
84. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with deep learning. Sci Rep 2018 Mar 15;8(1):4165 [FREE Full text] [doi: 10.1038/s41598-018-22437-z] [Medline: 29545529]
85. Singh VK, Romani S, Rashwan HA, Akram F, Pandey N, Sarker M. Conditional Generative Adversarial and Convolutional Networks for X-ray Breast Mass Segmentation and Shape Classification. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention. 2018 Presented at: MICCAI'18; September 16-20, 2018; Granada, Spain p. 833-840.
86. Agarwal V, Carson C. Stanford University. 2015. Using Deep Convolutional Neural Networks to Predict Semantic Features of Lesions in Mammograms URL: http://cs231n.stanford.edu/reports/2015/pdfs/vibhua_final_report.pdf
87. Gao F, Wu T, Li J, Zheng B, Ruan L, Shang D, et al. SD-CNN: a shallow-deep CNN for improved breast cancer diagnosis. Comput Med Imaging Graph 2018 Dec;70:53-62. [doi: 10.1016/j.compmedimag.2018.09.004] [Medline: 30292910]
88. Hagos YB, Mérida AG, Teuwen J. Improving Breast Cancer Detection Using Symmetry Information with Deep Learning. In: Proceedings of the Image Analysis for Moving Organ, Breast, and Thoracic Images. 2018 Presented at: RAMBO'18; September 16, 2018; Granada, Spain p. 90-97.
89. Teuwen J, van de Leemput S, Gubern-Mérida A, Rodriguez-Ruiz A, Mann R, Bejnordi B. Soft Tissue Lesion Detection in Mammography Using Deep Neural Networks for Object Detection. In: Proceedings of the 1st Conference on Medical Imaging with Deep Learning. 2018 Presented at: MIDL'18; July 4-6, 2018; Amsterdam, The Netherlands p. 1-9.
90. Jung H, Kim B, Lee I, Yoo M, Lee J, Ham S, et al. Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. PLoS One 2018;13(9):e0203355 [FREE Full text] [doi: 10.1371/journal.pone.0203355] [Medline: 30226841]
91. Shen L, Laurie LM, Joseph HR, Eugene F, Russell B, Weiva S. Cornell University. 2017. End-to-End Training for Whole Image Breast Cancer Diagnosis Using an All Convolutional Design URL: https://arxiv.org/pdf/1708.09427.pdf
92. GitHub. ISBI2018: Adversarial Deep Structural Networks for Mammographic Mass Segmentation URL: https://github.com/wentaozhu/adversarial-deep-structural-networks
93. [Department of Radiology]. URL: http://semmelweis.hu/radiologia/
94. GitHub. Computer aided detection using Faster-RCNN URL: https://github.com/riblidezso/frcnn_cad
95. GitHub. MAMMO_Retinanet URL: https://github.com/hwejin23/MAMMO_Retinanet
96. GitHub. End-to-end training for breast cancer diagnosis using deep all convolutional networks URL: https://github.com/lishen/end2end-all-conv

97. Lévy D, Jain A. Cornell University. 2016. Breast Mass Classification From Mammograms Using Deep Convolutional Neural Networks URL: https://arxiv.org/pdf/1612.00542.pdf
98. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Richter CD, Cha KH. Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. IEEE Trans Med Imaging 2019 Mar;38(3):686-696. [doi: 10.1109/TMI.2018.2870343]
99. Jadoon MM, Zhang Q, Haq IU, Butt S, Jadoon A. Three-class mammogram classification based on descriptive CNN features. Biomed Res Int 2017;2017:3640901 [FREE Full text] [doi: 10.1155/2017/3640901] [Medline: 28191461]
100. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imaging (Bellingham) 2016 Jul;3(3):034501 [FREE Full text] [doi: 10.1117/1.JMI.3.3.034501] [Medline: 27610399]
101. Domingues I, Cardoso J. Mass Detection on Mammogram Images: A First Assessment of Deep Learning Techniques. In: Proceedings of the 19th edition of the Portuguese Conference on Pattern Recognition. 2013 Presented at: RecPad'13; November 1, 2013; Lisbon, Portugal p. 2.
102. Wu E, Wu K, Cox D, Lotter W. Conditional Infilling GANs for Data Augmentation in Mammogram Classification. In: Proceedings of the Image Analysis for Moving Organ, Breast, and Thoracic Images. 2018 Presented at: RAMBO'18; September 16-20, 2018; Granada, Spain p. 98-106. [doi: 10.1007/978-3-030-00946-5_11]
103. Aboutalib SS, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S. Deep learning to distinguish recalled but benign mammography images in breast cancer screening. Clin Cancer Res 2018 Dec 1;24(23):5902-5909. [doi: 10.1158/1078-0432.CCR-18-1115] [Medline: 30309858]
104. Wang H, Feng J, Zhang Z, Su H, Cui L, He H, et al. Breast mass classification via deeply integrating the contextual information from multi-view data. Pattern Recognit 2018 Aug;80:42-52 [FREE Full text] [doi: 10.1016/j.patcog.2018.02.026]
105. Shams S, Platania R, Zhang J, Kim J, Lee K, Park SJ. Deep Generative Breast Cancer Screening and Diagnosis. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention. 2018 Presented at: MICCAI'18; September 16-20, 2018; Granada, Spain p. 859-867.
106. Gastounioti A, Oustimov A, Hsieh MK, Pantalone L, Conant EF, Kontos D. Using convolutional neural networks for enhanced capture of breast parenchymal complexity patterns associated with breast cancer risk. Acad Radiol 2018 Dec;25(8):977-984. [doi: 10.1016/j.acra.2017.12.025] [Medline: 29395798]
107. Dhungel N, Carneiro G, Bradley AP. Fully Automated Classification of Mammograms Using Deep Residual Neural Networks. In: Proceedings of the 14th International Symposium on Biomedical Imaging. 2017 Presented at: IEEE'17; April 18-21, 2017; Melbourne, Australia p. 310-314. [doi: 10.1109/ISBI.2017.7950526]
108. IEEE Xplore Digital Library. Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets URL: https://ieeexplore.ieee.org/document/8466816/media#media
109. Department of Computer Science, University of Toronto. Training a deep autoencoder or a classifier on MNIST digits URL: http://www.cs.toronto.edu/~hinton/MatlabForSciencePaper.html
110. GitHub. Code for "Conditional infilling GAN for Data Augmentation in Mammogram Classification" URL: https://github.com/ericwu09/mammo-cigan


## Abbreviations

**BI-RADS:** Breast Imaging Reporting and Data System
**CAD:** computer-aided diagnosis
**CC:** craniocaudal
**CEDM:** contrast-enhanced digital mammograms
**cGAN:** conditional generative adversarial network
**CNN:** convolutional neural network
**CRF:** conditional random field
**CT:** curvelet transform
**DBT:** digital breast tomosynthesis
**DiaGRAM:** Deep Generative Multitask
**DL:** deep learning
**DM:** digital mammogram
**DW:** discrete wavelet
**FCN:** fully convolutional network
**FP:** false positive
**IARC:** International Agency for Cancer Research
**KNN:** k-nearest neighbor
**LDA:** linear discriminant analysis
**LE:** low energy
**ML:** machine learning

**MLO:** mediolateral oblique
**PM:** pectoral muscle
**ReLU:** rectified linear unit
**ResNet:** residual network
**ROI:** region of interest
**SAE:** stacked autoencoder
**SD-CNN:** shallow-deep convolutional neural network
**SVM:** support vector machine
**TRW:** tree reweighted
**VGG:** visual geometry group
**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 23.04.19; peer-reviewed by M Awais, E Javed, M Hamghlam; comments to author 30.05.19; revised version received 11.06.19; accepted 12.06.19; published 26.07.19.*

*Please cite as:*
*Gardezi SJS, Elazab A, Lei B, Wang T*
*Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review*
*J Med Internet Res 2019;21(7):e14464*
*URL: http://www.jmir.org/2019/7/e14464/*
*doi:10.2196/14464*
*PMID:31350843*

Original Paper

# Cloud Health Resource Sharing Based on Consensus-Oriented Blockchain Technology: Case Study on a Breast Tumor Diagnosis Service

Xiaobao Zhu[1,2], MSc; Jing Shi[2], PhD; Cuiyuan Lu[2], MSc

[1]School of Software, Nanchang Hangkong University, Nanchang, China

[2]Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, OH, United States

**Corresponding Author:**
Jing Shi, PhD
Department of Mechanical and Materials Engineering
University of Cincinnati
2600 Clifton Ave
Cincinnati, OH,
United States
Phone: 1 513 556 2380
Fax: 1 513 556 2380
Email: jing.shi@uc.edu

## Abstract

**Background:** In recent years, researchers have made significant efforts in advancing blockchain technology. This technology, with distinct features of decentralization and security, can be applied to many fields. In areas of health data and resource sharing, applications of blockchain technology are also emerging.

**Objective:** In this study, we propose a cloud health resource-sharing model based on consensus-oriented blockchain technology and have developed a simulation study on breast tumor diagnosis.

**Methods:** The proposed platform is built on a consortium or federated blockchain that possesses features of both centralization and decentralization. The consensus mechanisms generate operating standards for the proposed model. Open source Ethereum code is employed to provide the blockchain environment. Proof of Authority is selected as the consensus algorithm of block generation.

**Results:** Based on the proposed model, a simulation case study for breast tumor classification is constructed. The simulation includes 9893 service requests from 100 users; 22 service providers are equipped with 22 different classification methods. Each request is fulfilled by a service provider recommended by the weighted k-nearest neighbors (KNN) algorithm. The majority of service requests are handled by 9 providers, and provider service evaluation scores tend to stabilize. Also, user priority on KNN weights significantly affects the system operation outcome.

**Conclusions:** The proposed model is feasible based on the simulation case study for the cloud service of breast tumor diagnosis and has the potential to be applied to other applications.

***(J Med Internet Res 2019;21(7):e13767)*** doi:10.2196/13767

**KEYWORDS**

blockchain; cloud health; breast tumor diagnosis; k-nearest neighbors (KNN); Proof of Authority (PoA); consensus-oriented

## Introduction

### Background

Health care is closely related to the survival and happiness of human beings, and thus the efficiency and effectiveness of health care are of critical importance. The health care industry is one of the most important industries for developed and developing countries. According to a report from the World Health Organization, in 2018 total health care expenditure grew by 4% for high-income countries, while this figure reached approximately 6% for low- and middle-income countries. In either group of countries, the growth rate of health care expenditure is higher than that of gross domestic product [1].

In recent decades, information and computer technologies have significantly improved the efficiency of health care delivery and access to care and greatly reduced the waste of health care resources. Besides the well-known examples of electronic health records (EHRs), telemedicine, and clinical decision support systems, applications of new technologies such as mobile health and artificial intelligence (AI) are booming. For instance, HealthTap [2], a popular health app, offers health services at no additional charge to policyholders that include asking a network of licensed physicians health-related questions online by connecting immediately or by appointment with a doctor for consultation via video conference, phone call, or text chat. It has now attracted more than 140,000 licensed doctors in good standing from 170 countries. Also, Google's DeepMind Health [3] is a health AI system collaborating with the National Health Service in the United Kingdom. The goal is to deliver better care for millions of people worldwide.

However, a rising number of issues have been reported along with the digitalization and informatization in health care fields. It is well recognized that information systems can fail to deliver the best solution for the patients due to the lack of necessary information [4]. Adler-Milstein et al [5] illustrated that customized and incompatible health systems can cause gaps in communication and coordination between medical organizations. Zhang et al [6] believed that in health information systems, one fundamental problem is the lack of a trusted platform that can connect independent health systems and provide an end-to-end reachable network. Similarly, Zhang et al [7] also indicated that pressing issues in the health field include fragmented and siloed data, delayed communications, disparate workflow tools, and the lack of a health care resource–sharing platform. In this regard, blockchain technology, with its unique characteristics such as decentralization, consensus, cryptocurrency, and immutability of data, provides a novel tool to address these issues.

In this study, we propose a cloud health resource–sharing model based on consensus-oriented blockchain technology and illustrate the model with a case study on breast tumor diagnosis.

## Literature Review

### *Blockchain in Health Data Sharing*

Health data sharing has been one of the biggest challenges for health care organizations. Since Bitcoin was first introduced in 2008 by the pseudonymous creator Satoshi Nakamoto [8], it has experienced amazing development. Blockchain, which is the core technology of Bitcoin, has drawn unprecedented interest and attention. In the past several years, researchers and practitioners have started to recognize the value of blockchain technology for addressing data sharing challenges. Rifi et al [9] illustrated the specific problems such as privacy, scalability, and interoperability and highlighted the benefits of blockchain technology in the deployment of a secure and scalable solution for medical data exchange. Xia et al [10] addressed patient privacy risks of disseminating medical data beyond the protected cloud of institutions and proposed a blockchain-based data-sharing framework. The framework addresses access control challenges associated with sensitive data stored in the cloud using the immutability and built-in autonomy properties of blockchains. Liang et al [11] proposed an innovative user-centric health data-sharing solution by using a decentralized and permissioned blockchain for privacy protection and identity management improvement. More recently, Alexaki et al [12] also considered blockchain technology for supporting the decentralized care cycle. With blockchain technology, patient privacy and medical record integrity is addressed, while efficient interoperability between providers is simultaneously ensured. Zhang et al [5] illustrated the contributions of blockchain technology for clinical data sharing in the context of technical requirements defined in the Shared Nationwide Interoperability Roadmap from the Office of the National Coordinator for Health Information Technology. In addition, Ji et al [13] proposed a location-sharing scheme for telecare medical information systems by using blockchain technology. In their work, basic requirements of the scheme such as decentralization, confidentiality, and verifiability were defined and an experiment was conducted to demonstrate the efficiency and feasibility of the proposed scheme.

### *Blockchain in Electronic Health Records*

In literature, numerous studies have applied blockchain technology to EHR management. Dey et al [14] developed a solution of reliable storage of health data by proposing a blockchain–Internet of Things model where a biosensor measures and collects real-time data concerning a patient's medical status and stores the data in the blockchain. The InterPlanetary File System protocol was proposed to save discharged patient records, thus reducing the load on the actual blockchain. Li et al [15] proposed a novel blockchain-based data preservation system for medical data in which users can preserve the essential data and the originality of data. A prototype of the data preservation system was implemented based on the real-world blockchain-based platform Ethereum, and the results demonstrated the effectiveness and efficiency of the proposed system. Dagher et al [16] proposed a blockchain-based framework called Ancile. The Ancile framework can provide secure, interoperable, and efficient access to medical records for patients, providers, and third parties while preserving the privacy of patient-sensitive information. Chen et al [17] designed a storage scheme to manage personal medical data based on blockchains and cloud storage. Furthermore, a service framework for sharing medical records was described. Also, the characteristics of the medical blockchain were presented and analyzed through comparison with traditional systems. Wang and Song [18] proposed a secure EHR system with an attribute-based cryptosystem for medical data, identity-based encryption for digital signatures, and blockchain technology for the integrity and traceability of medical data. Similarly, Guo et al [19] indicated there is a critical need for patients to pay close attention to their own health care information and medical data storage. An attribute-based signature scheme with multiple permissions was proposed to ensure the effectiveness of EHR infused in the blockchain.

Zhang et al [20] described the issues of system evolvability, storage requirements minimization, patient data privacy protection, and application scalability across a large number of users. These challenges can be mitigated in a blockchain-based decentralized application (DApp) for smart health. Brogan et al [21] discussed how distributed ledger technologies can play a key role in advancing electronic health by ensuring the authenticity and integrity of data generated by wearable and embedded devices. Tian et al [22] proposed to establish a shared key that could be reconstructed by the legitimate parties before the process of diagnosis and treatment begins. The data in the diagnosis and treatment process are encrypted and stored in a blockchain using the shared key.

### *Blockchain in Medicine Prescription Tracking*

Blockchain technology provides the health industry a new vision for drug tracking, in particular opioid prescription tracking. Mettler et al [23] demonstrated the examples of public health care management, user-oriented medical research, and drug counterfeiting in the pharmaceutical sector. The examples were believed to be just the starting points for blockchain technology to be adopted in the health care industry. Dhillon et al [24] proposed a blockchain system in which a provider can check the blockchain to find a currently active prescription when writing a prescription. An active prescription from a different provider will automatically invalidate the request for a new prescription, and this can be encoded as a second spending request in the network. Meanwhile, efforts have been made by major blockchain participants such as IBM and Deloitte blockchain laboratories to control opioid overdose epidemics. Zhang et al [5] indicated that blockchain-based systems can build a trusted network of hospitals and pharmacies to store drug-related transactions in a responsible way. The distributed and shared-licensed blockchain platform allows loosely coupled providers to access other data silos without a clear trust relationship among them. Taylor and Hare [25] employed the permissions and restrictions associated with the digital wallet to interact with unexpected events and requirements of transactions contained in blocks. This interaction can be used to realize the verification of opioid dose ownership. It may also include provisions for the sale of opioids that involve current owners, patients, and drug abuse agencies.

### *Blockchain in Clinical Trials and Precision Medicine*

The online service of clinical trials and precision medicine is becoming increasingly popular. Blockchain technology provides a trustworthy safety mechanism for this service. Shae and Tsai [26] proposed a blockchain platform for clinical trials and precision medicine, and they identified four new system architecture components required to be developed on top of the traditional blockchain. Suzuki et al [27] proposed a scheme to record both client requests and server replies in an auditable manner using blockchain technology as a request-response channel for a client-server system. A proof-of-concept algorithm was developed based on a publicly available blockchain testbed. Tsai [28] proposed a mechanism for transforming repetitive blockchain calculations into a distributed parallel computing architecture. In the process, smart contracts are adopted to support mobile computing. The mechanism is elucidated to establish real-world evidence of clinical trials for individuals and precision medicine. Benchoufi et al [29] adopted blockchain technology to build a consent workflow. The proposed proof-of-concept protocol includes the use of a blockchain to time-stamp each step in the patient's consent to collect clinical trial information in a secure, indivisible, and transparent manner through cryptographic verification. A single document is obtained in an open format that explains the entire consent collection process. It is believed that in the future, blockchains can be used to track complex data from clinical trials, and streaming smart contracts can help prevent clinical trial events from occurring in incorrect chronological order.

## Methods

### Concept of the Blockchain-Based Cloud Health Service-Sharing Model

The proposed system presents a new model of consensus-oriented health data and service sharing by integrating the blockchain technologies with the concept of cloud computing. Blockchain techniques such as public key and private key technology, virtual currency, smart contract, consensus algorithm, and HASH256 encrypted technology are used for automatic consensus-driven services and sharing of value. The services allow health care organizations, health platforms, individuals, and health-relevant industry to share the increasing system value and possess a variety of safe, reliable, credible, high-quality, inexpensive, easily payable, and on-demand health resources.

The proposed health service–sharing model is based on an open source system. All of the consensus standards and system recommendation algorithms are open source to the approved users. The proposed model adopts the consortium or federated blockchain structure [30]. Therefore, the blockchain system in the proposed model is not fully decentralized. Instead, the consensus-oriented centralized model helps the system to stay away from the potential issues of a fully decentralized system such as crime and volatility. Users in the proposed model are divided into three major categories: administrators, service providers, and regular users. Administrators typically include signers and members of arbitration committees. The information of service providers is pre-verified by an administrator, and the administrators vote through smart contracts. The blockchain system only stores transaction and summary information related to system components such as the basic machine specification. However, there are many types of information (eg, medical images) that need to be verified by the blockchain system in the proposed model. Those data are typically saved in a distributed storage system and verified by the Oracle mechanism. Oracle in a blockchain system provides trusted entities that allow the blockchain system to access external data [31]. The Oracle mechanism also guarantees the safety of external data blockchain data.

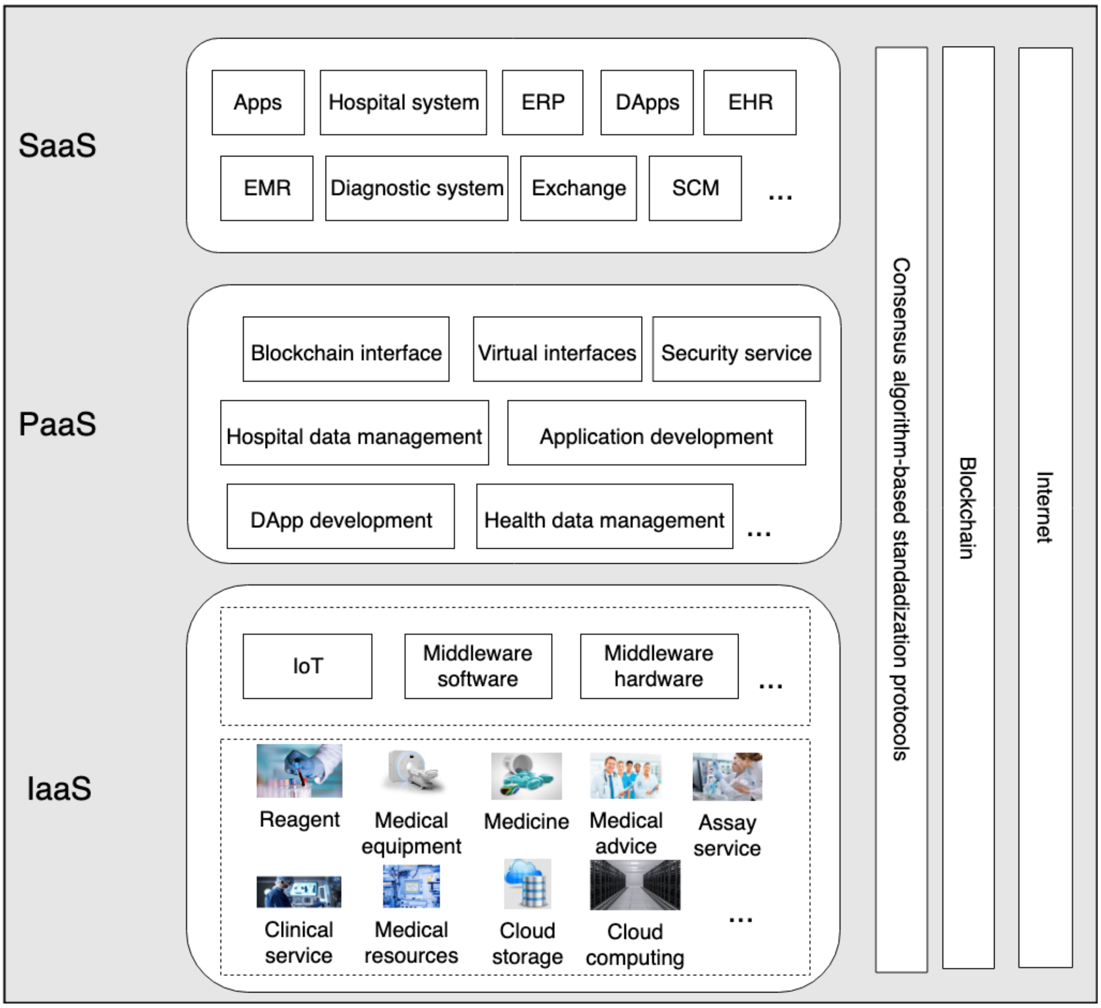
Value sharing is the core concept of the proposed model. The model employs a cryptocurrency system, the cloud health coin system, and the cryptocurrency is called cloud health coin (CHC). In the blockchain system of the proposed model, only signers can mine blocks. Therefore, Proof of Authority (PoA) [32] is employed as the network consensus algorithm rather than Proof of Work. Moreover, the practical Byzantine fault tolerance protocol can ensure system safety even with individual signer errors. With the PoA consensus algorithm, it is not necessary for signers to invest significant funds for computing power competition. Besides the conventional design in which signers are entitled to all the mining rewards, we provide another option in which a considerable proportion of mining rewards is transferred into a system public account. The public fund is used to support a standard verification smart contract, which rewards data sharing and those who have contributed to the system. For instance, the equipment that has spare computing power and is involved in the distributed computing of the recommending provider will receive a reward.

### Hierarchical Structure for Everything as a Service on Proposed Model

The concept of *everything as a service* is adopted from cloud computing and applied in the proposed model. Everything is virtualized to serve customers by applying standards of consensus mechanism through the smart contracts. Also, the system allows the health platform to use and share the health data comfortably and conveniently. Physical cloud health resources deliver *infrastructure as a service* by core middleware capabilities. The user level middleware and system developer tools provide *platform as a service* capabilities. The various application services, DApps, and professional software offer *software as a service* capabilities. Figure 1 shows the hierarchical structure for *everything as a service* on the proposed model.

The *software as a service* layer provides the services of many software applications to the end users. The software can be provided by the proposed system or third-party companies or even developed by users. The majority of the software packages are not open source. Examples of software applications include hospital management systems, EHRs, electronic medical records, and CHC exchange. The *platform as a service* layer offers software frameworks that help developers create apps, DApps, or other *software as a service* layer software. This level also provides a number of open source tools preapproved by smart contracts. These tools allow developers to develop apps or DApps according to relevant system standards comfortably. All health care organizations provide the interface of standardized data access (with consensus). Middleware software enables the data from various platforms to have a standard format. The *platform as a service* layer also includes service level agreements, accounting, billing, Oracle, and blockchain system interfaces.

**Figure 1.** Hierarchical structure for everything as a service for the proposed model. SaaS: software as a service; PaaS: platform as a service; IaaS: infrastructure as a service; ERP: enterprise resource planning; DApp: decentralized application; EHR: electronic health record; EMR: electronic medical record; SCM: supply chain management; IoT: internet of things.



The *infrastructure as a service* layer includes medical systems for diagnosis and treatment, laboratory equipment, computers, and other health resources that can be made available on the cloud. In the proposed model, the service providers offer services according to consensus standards. The middleware helps service providers deliver services through a virtualization process.

### Case Study on Cloud-Based Breast Tumor Diagnosis

Assume that a breast cancer patient who lives in a developing country or rural area needs laboratory examination of her biological sample to identify the cancer type. Hospitals in her country or area cannot identify the biological sample, so the patient must resort to hospitals or laboratories in other countries. The challenges now: How can the patient find an appropriate organization for the diagnosis in another country? How can the patient schedule services from the identified organization? How can the patient pay for the services?

With the proposed model, those challenges faced by the patient can be addressed. First, all service providers in the system are preapproved. The system uses open source algorithms to recommend providers who are qualified. The nature of open source algorithms ensures impartial and more credible recommendation results compared with the results of an internet search using commercial search engines. Virtual currency is applied to pay for the transaction, which frees the patient from the traditional currency exchange. The patient can simply submit a service request, often with an affordable cost limit, to the system and wait for the notification. As shown in Figure 2, the system responds to a service request initiated by the patient (user). The system will provide the estimated cost range based on a big data analysis of past similar services and can provide a convenient service to buy CHC. Meanwhile, the system will select a signer to organize the computing process for recommending a service provider according to consensus mechanism. If the concept of a public account is adopted, a number of users and providers can be involved in distributed computing and obtain their computing rewards. This can enable users and providers who are not signers to share the system value by dividing up the gas and mining rewards with the signer. The providers related to this service request receive the detailed job requirement and payment information. All process data are written onto the blockchain for retrospective purposes.

Thereafter, patients/users will have the test sample prepared according to the system standard. When the test sample is ready, the processes will start and all related processes are monitored under open source algorithms. After each process is finished, the patient can rate the services. The result of the evaluation will be written to the blockchain. If a provider does not agree with the assessment by the patient, an application of arbitration can be submitted. All shreds of evidence guarantee authenticity and credibility of the testimonies. During the process, a vote will happen in the arbitration committee. This vote relies on data reports and is realized by the smart contract.

**Figure 2.** Breast tumor diagnosis based on the proposed model. DApp: decentralized application.

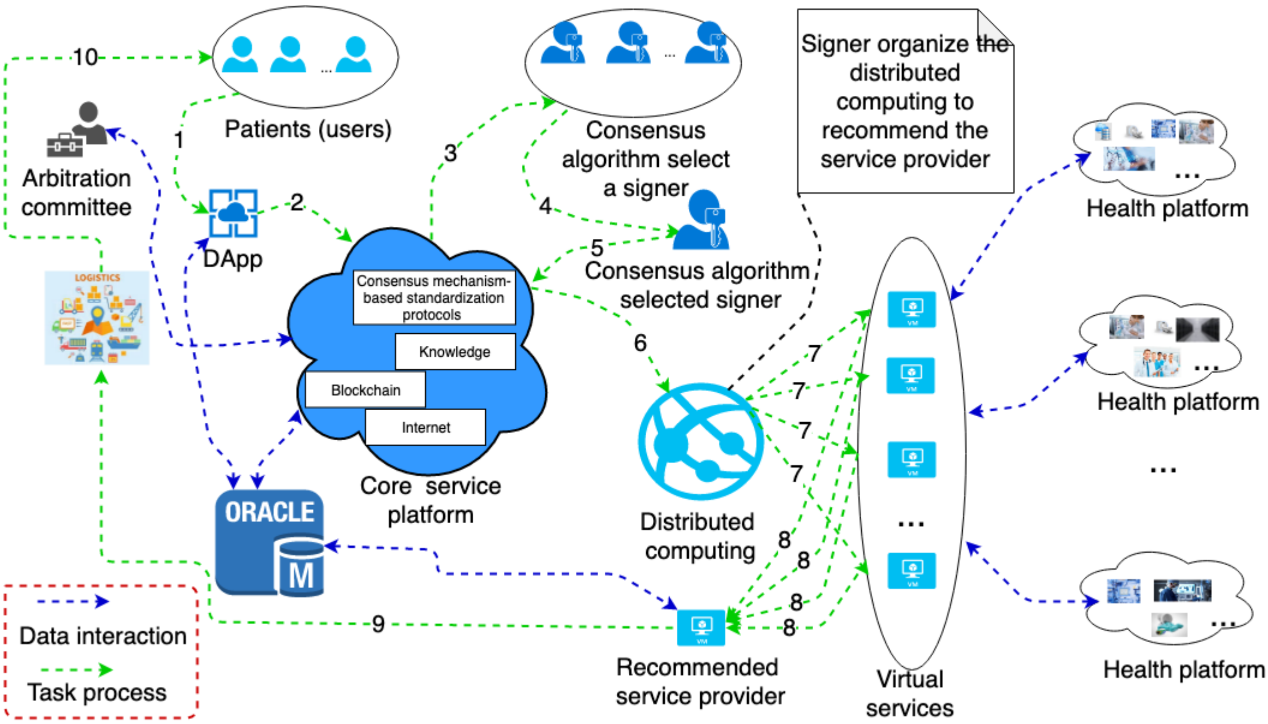
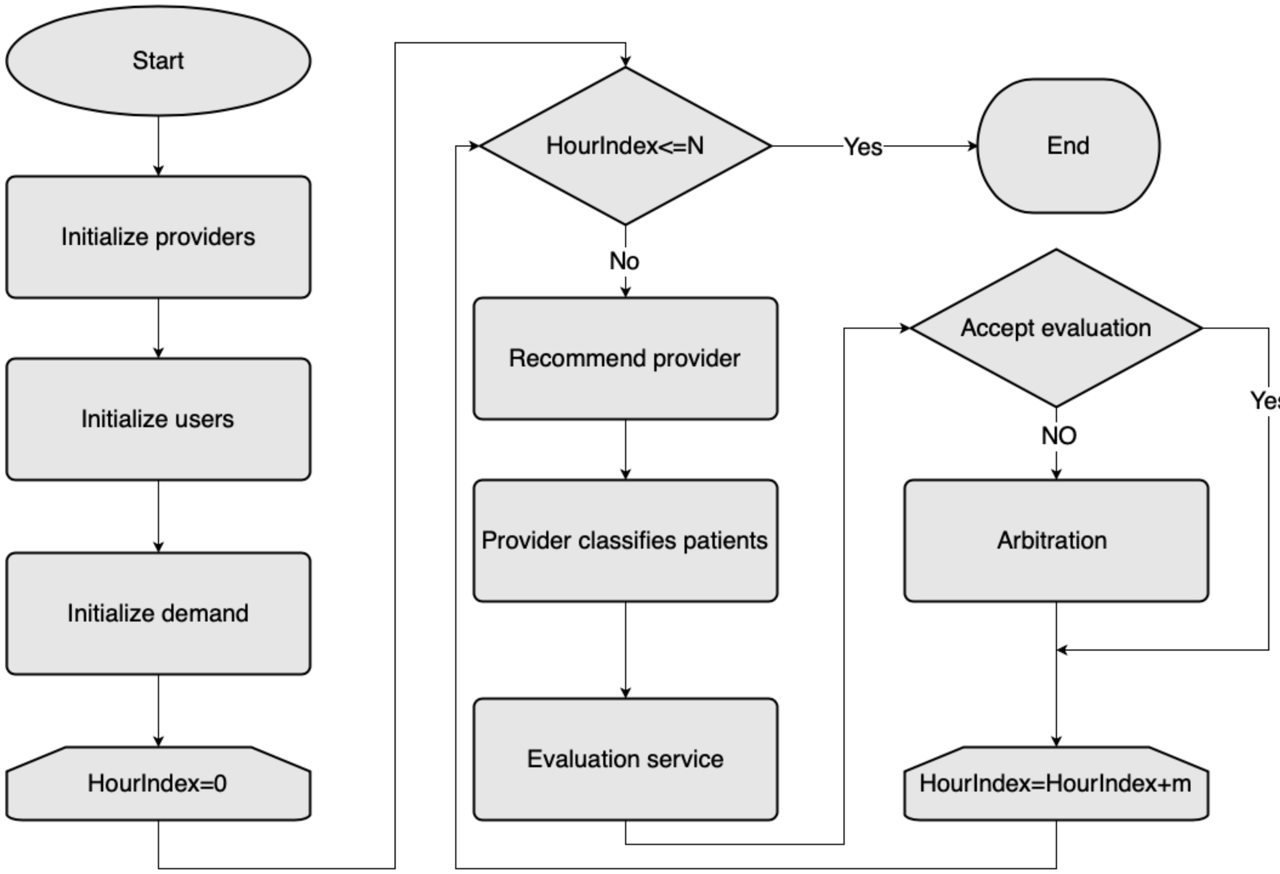


**Figure 3.** Simulation flowchart.




## Simulation

We implemented the proposed model based on the open source blockchain system Ethereum. The simulation was coded in Python 3.6 with a Spyder integrated development environment carried out on a Mac system with a 2.2 GHz Intel i7 CPU and 16 GB DDR4 memory. Three computers running macOS, Ubuntu, and Windows operating systems were employed to simulate the cross-platform scenario with multiple operating systems. We constructed a consortium or federated blockchain based on Ethereum. The system assumes 9 predetermined signers (also can be called owners), who are the only accounts that can mine the blocks according to the PoA consensus algorithm, and 22 providers who can provide the breast cancer diagnosis service. One hundred users are created to make service requests. A Poisson distributed random function is used to simulate the interarrival time of job requests. The purpose of the simulation is to gain insights into the performance and evaluation of providers. Figure 3 shows the simulation flowchart, which includes initiation, provider recommendation, and service evaluation.

## Provider Estimation

Based on the Wisconsin Diagnosis Breast Cancer Database [33,34], researchers have developed methods for breast cancer data classification. Table 1 summarizes the methods and their accuracies of classification according to the literature. In this study, each method is treated as a provider.

Considering smart contracts have preapproved all providers, the provider's initial service evaluation score is set to be 6 out of 10. In the 0 to 10 evaluation scale, 0 represents the worst, while 10 represents the best. A patient document is regarded as a data unit. Providers have different pricing strategies on the unit price. Here, we assume that a higher accuracy requirement leads to a more expensive diagnosis cost. The unit price ($p_{unit}$) is calculated by $p_{unit} = \delta + \varepsilon * f(\alpha)$, in which $\delta$ and $\varepsilon$ are constants, $\delta$ represents the base price, $\varepsilon$ represents the price variance caused by the accuracy rate of diagnosis, and $f(\alpha)$ is a normal distribution random-based function that ranges from 0 to 1. The mean of $f(\alpha)$ equals $\alpha$, which is the normalized diagnostic accuracy.

The diagnostic accuracy has a stochastic nature associated with the number of jobs serviced. The initial diagnosis is defined by Table 1. The simulation has a computer diagnosis function. The diagnostic accuracy of each provider will change as the number of jobs fulfilled increases. The dynamic diagnostic accuracy is calculated as seen in Figure 4, in which $\gamma$ is a constant to prevent an accidental diagnosis that affects the provider's accuracy too much. In this case, $\gamma$ equals 1000, $\alpha_{ini}$ is the initial diagnostic accuracy of a provider as shown in Table 1, and $\rho$ is the total number of serviced cases. As seen in Figure 5, $\alpha_{updated}$ is calculated where $\beta_{accuracy}$ represents the number of patient cases diagnosed accurately by the provider and $\beta_{total}$ represents the overall number of patient cases diagnosed by the provider.

**Table 1.** Breast cancer diagnosis methods (providers).

| Provider # | Method | Accuracy, % |
|---|---|---|
| 1 | CfS[a] + SVM[b] [35] | 87.84 |
| 2 | Filtered + SVM [35] | 87.84 |
| 3 | CfS + logistic regression [35] | 95.95 |
| 4 | Filtered + logistic regression [35] | 96.62 |
| 5 | BPSO[c]-2Stage [36] | 92.98 |
| 6 | PSO[d] (4-2) [36] | 93.98 |
| 7 | KP[e]-SVM [37] | 97.55 |
| 8 | RFE[f]-SVM [37] | 95.25 |
| 9 | FSV[g] [37] | 95.23 |
| 10 | Fisher + SVM [38] | 94.70 |
| 11 | Self-training [38] | 85.12 |
| 12 | Random co-training [38] | 83.54 |
| 13 | Rough co-training [38] | 88.63 |
| 14 | LDA[h] [39] | 97.19 |
| 15 | C4.5 [39] | 94.06 |
| 16 | DIMLP[i] [39] | 96.92 |
| 17 | SIM[j] [39] | 98.26 |
| 18 | MLP[k] [39] | 97.43 |
| 19 | PSO-KDE[l] (1) [40] | 98.45 |
| 20 | PSO-KDE (2) [40] | 98.45 |
| 21 | GA[m]-KDE (2) [40] | 98.45 |
| 22 | Fisher + PFree Bat[n] + LS[o]-SVM [41] | 100 |

[a]CfS: correlation-based feature selection.
[b]SVM: support vector machine.
[c]BPSO: binary particle swarm optimization.
[d]PSO: particle swarm optimization.
[e]KP: kernel-penalized SVM (KP-SVM).
[f]RFE: recursive feature elimination.
[g]FSV: feature selection concave.
[h]LDA: linear discriminant analysis.
[i]DIMLP: discretized interpretable multilayer perceptron.
[j]SIM: similarity classifier.
[k]MLP: multilayer perceptron.
[l]KDE: kernel density estimation.
[m]GA: genetic algorithm.
[n]PFree Bat: parameter-free bat optimization algorithm.
[o]LS: least square support vector machine.

**Figure 4.** Equation of computing dynamic diagnostic accuracy.

$$\alpha = \frac{\gamma * \alpha_{ini} + \rho * \alpha_{updated}}{\gamma + \rho}$$

**Figure 5.** Equation of computing historical dynamic diagnostic accuracy.

$$\alpha_{updated} = \frac{\beta_{accuracy}}{\beta_{total}}$$

## Modeling the Service Requests

In the proposed system, everything is virtualized as a service according to the consensus standards. In this simulation, a computing service for breast tumor classification is completed. A Poisson random function is employed to simulate an interarrival time of generated service requests. The average interarrival time of service request is 3 minutes. A total of 9893 job requests are generated during 500 hours of the simulation run. Each service request includes a dataset containing a varying number of images to be analyzed. User priorities in terms of cost sensitivity and diagnostic accuracy sensitivity are reflected by the k-nearest neighbors (KNN) weights. Therefore, different users may have different KNN weight combinations.

## Design and Implementation of Breast Tumor Diagnosis Services

The price quote ($p_{quote}$) for the job is calculated by $p_{quote}= \beta * p_{unit}$, in which β is the number of patient cases in the job and $p_{unit}$ is calculated by $p_{unit}= \delta + \varepsilon * f(\alpha)$.

Figure 6 shows the recommendation process based on the weighted KNN algorithm [42]. In this case, the total cost, diagnostic accuracy, and score of service evaluation are considered as the three nearest neighbors in the KNN algorithm. To estimate the different preferences of users, users have specific KNN weights. Based on KNN weight characteristics, users can be classified into four types. Type 1 users are price-oriented, and the weight of the total price is equal to or greater than 0.5; type 2 users are accuracy-oriented, and the weight of diagnostic accuracy is equal to or greater than 0.5; type 3 users are reputation-oriented, and the weight of the service evaluation score is equal to or greater than 0.5; and type 4 users are the normal users who do not have a single weight equal to or greater than 0.5.

Peterson et al [43] indicate that the majority of conceptual distributions of satisfaction measurements follow a skewed distribution. As a result, a skewed distribution function is employed to model user evaluation results. Providers who have not successfully bid for a job will receive the average system service evaluation score. The equation in Figure 7 is used to compute the score of service evaluation of providers, where n represents the number of jobs that have been completed by the provider, e represents the evaluation score e∈[0,10], and $\phi_i$ is the evaluation coefficient of user, $\phi_i \in \{0,0.2,0.4,0.6,0.8,1\}$. The symbol ε represents the difference between the user evaluation and the provider self-appraisal, which is evaluated by the equation in Figure 8, in which $E_i^{self}$ represents the provider self-evaluation regarding the service. Finally, the value of $\phi_i$ is computed by the equation in Figure 9.

To simulate the process of arbitration, the difference between the provider self-appraisal and the user rating is calculated. The values of the difference are classified into several categories according to the probabilities as shown in Figure 10. A uniform distribution random value is generated to compare the values of probability. If the random value is less than the probability, arbitration is triggered, and vice versa.

**Figure 6.** Process flow to obtain recommended provider. KNN: k-nearest neighbors.

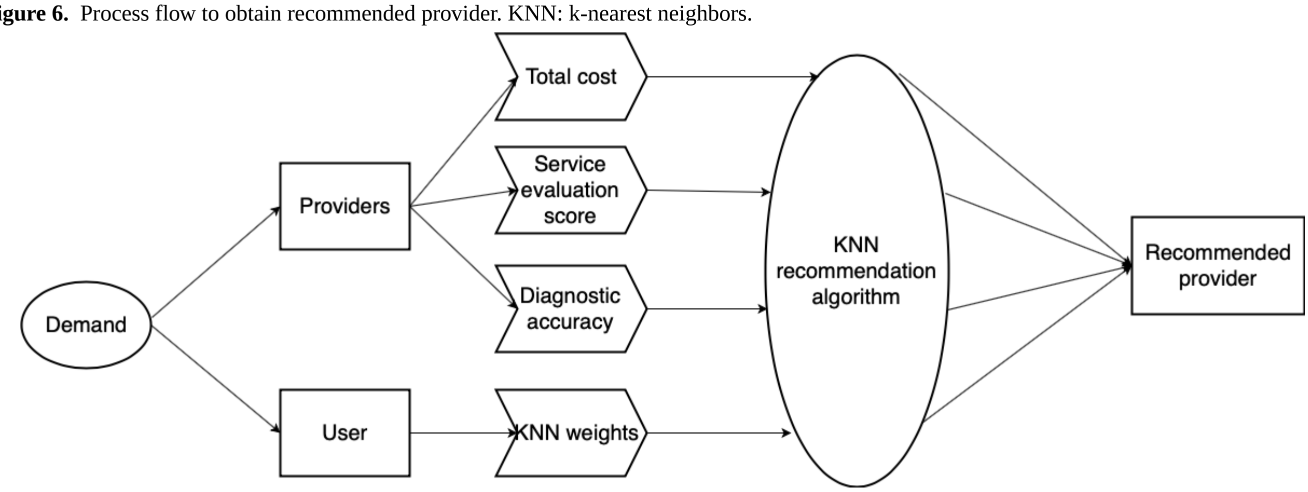



**Figure 7.** Equation of computing the service evaluation score.

$$E_i = \frac{\sum_i^n e_i * \varphi_i}{\sum_{i=0}^n \varphi_i}$$

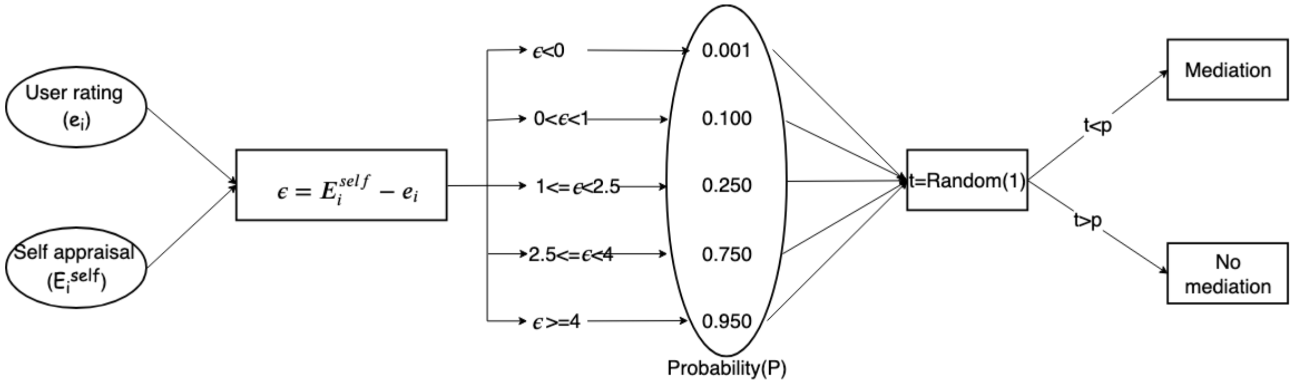
**Figure 8.** Equation of computing the difference between the user evaluation and the provider self-appraisal.

$$\varepsilon = \left|e_i - E_i^{self}\right|$$

**Figure 9.** Equation of computing the probability of an arbitration request.

$$\varphi_i = Fun(\varepsilon) = \begin{cases} 1, & 0 \le \varepsilon \le 1 \\ 0.8, & 1 < \varepsilon \le 2 \\ 0.6, & 2 < \varepsilon \le 3 \\ 0.4, & 3 < \varepsilon \le 4 \\ 0.2, & 4 < \varepsilon \le 9 \\ 0, & 9 < \varepsilon \le 10 \end{cases}$$

**Figure 10.** Process of determining arbitration for user ratings of service.




## Results

In the simulation, 500 hours of interarrival time of service request generation is simulated and 9893 service requests are generated. It is found that 15 providers have fulfilled all 9893 requests, while the other 7 providers have failed to take up any job. Figure 11 shows the appraisal of provider service given by users. The figure indicates that service evaluation scores of each provider stabilize with the increase of job index. This is because the service evaluation of users is related to the diagnostic accuracy, listed in Table 1. The figure shows rapid increases in evaluation score in the initial phase. This is because the service level scores of providers are set to 0 at the beginning, with scores updated after the providers start to fulfill service requests. However, for providers who fail to fulfill any job, changes in service evaluation score can also be observed. The reason is that those providers are assigned the average service evaluation score of the system. This mechanism enables all providers to have a winning opportunity to compete for jobs. It can also be found that the curves possess two different shapes. Some curves appear to comprise several major completely flat line segments, while others show a wavy pattern. The first type of curve shape implies that a provider only takes up a very limited number of service request. Many providers, such as providers 7, 8, and 21, show a sudden drop right after a sudden increase around service request numbers 2400, 2073, and 600, respectively. Those providers fulfill no service requests before that. Once they take up their first service requests, their service evaluation changes from the system average to a very high evaluation score. However, after that, those providers fulfill approximately 10 service requests in a short amount of time and then the service score reduces and stabilizes.

Figure 12 illustrates the final total number of service requests fulfilled by each provider during the simulated 500 hours. It can be seen that provider 22 fulfills the most service requests, which comprise almost a quarter of the total requests. Providers 2, 9, and 10 fulfill more than 1000 service requests. Also, providers 1, 11, 12, and 19 represent the third echelon, with fulfilled requests higher than 500 but less than 1000. The 8 providers undertake 9254 out of 9853 service requests (more than 90%). Figure 13 shows the service score for the first 50 arbitration cases. It shows that the majority of mediation results are higher than the original user evaluations. The result implies that to a certain extent, the arbitration mechanism has the ability to fix the difference between the score given by the users and the actual service level of the provider. On the other hand, the arbitration results of cases 9, 13, 24, 29, 33, and 38 are very similar to the corresponding user evaluations. Moreover, it is evident that for case 8, the mediation result is actually lower than the user evaluation. The observations indicate that while the arbitration mechanism does offer the capability of fixing malicious reviews of the providers, it does not guarantee a better result than the user evaluation.

**Figure 11.** Service evaluation changes during simulation.

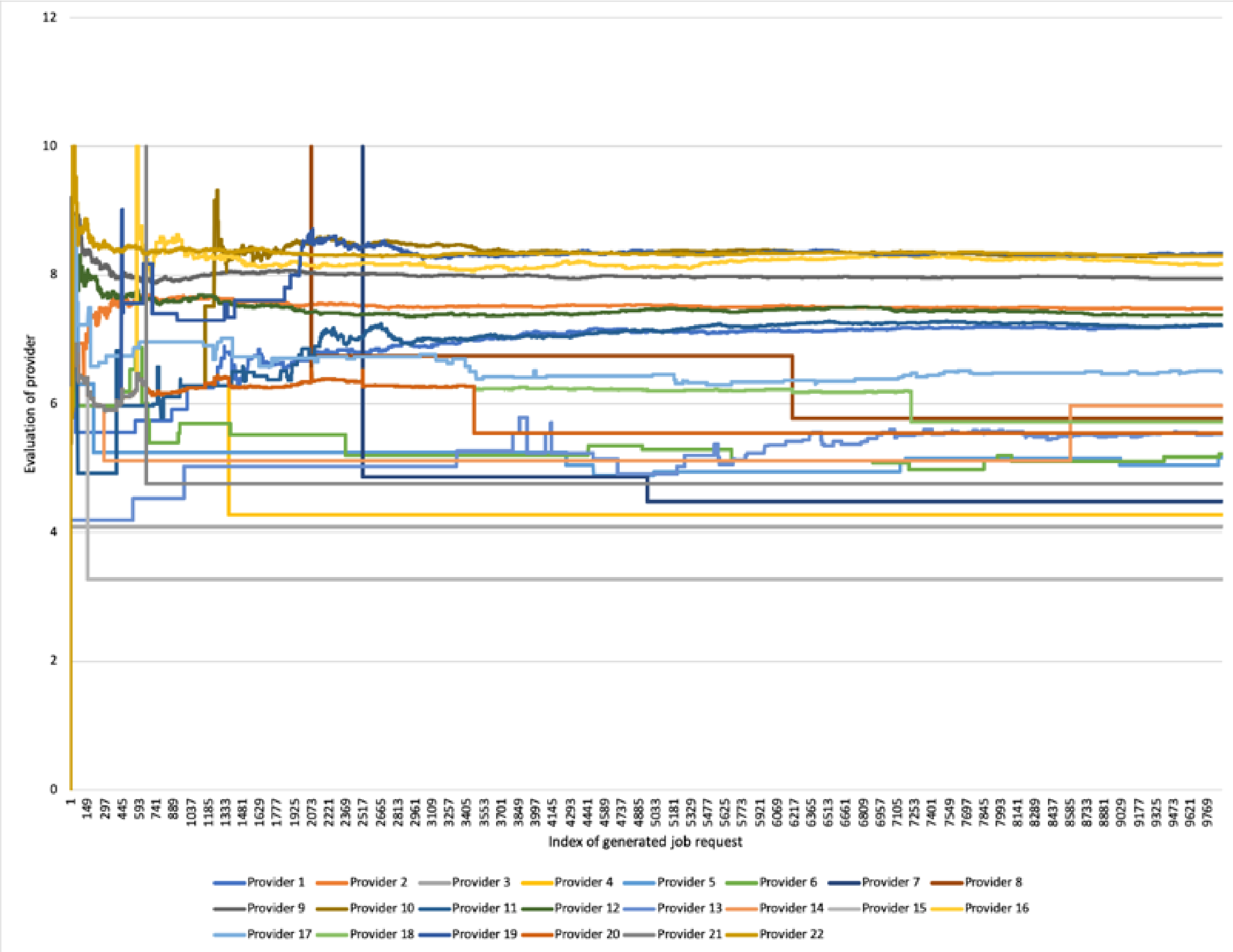


**Figure 12.** Number of service requests fulfilled by providers.

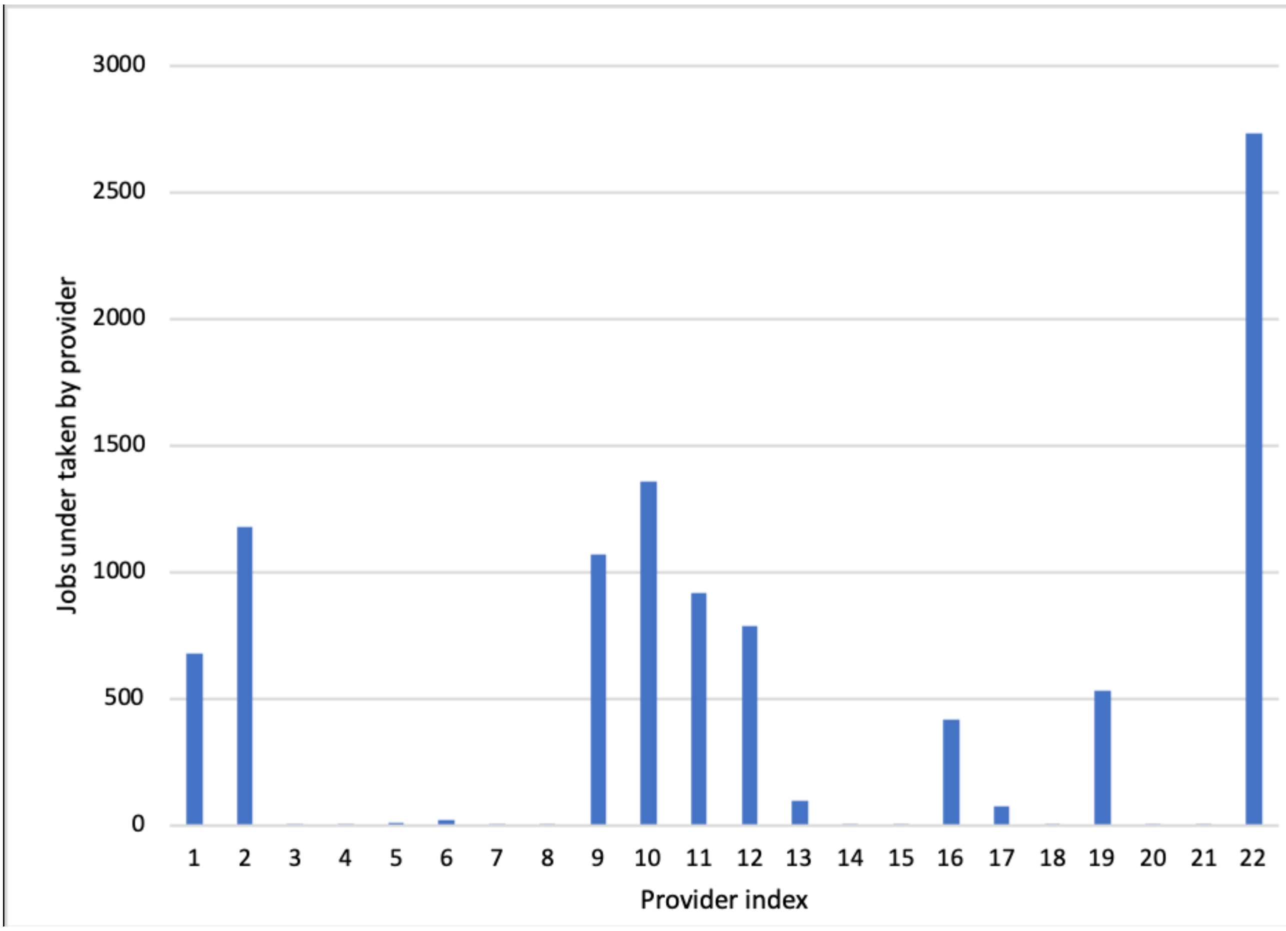



**Figure 13.** Evaluation of provider in first the 50 mediation cases.

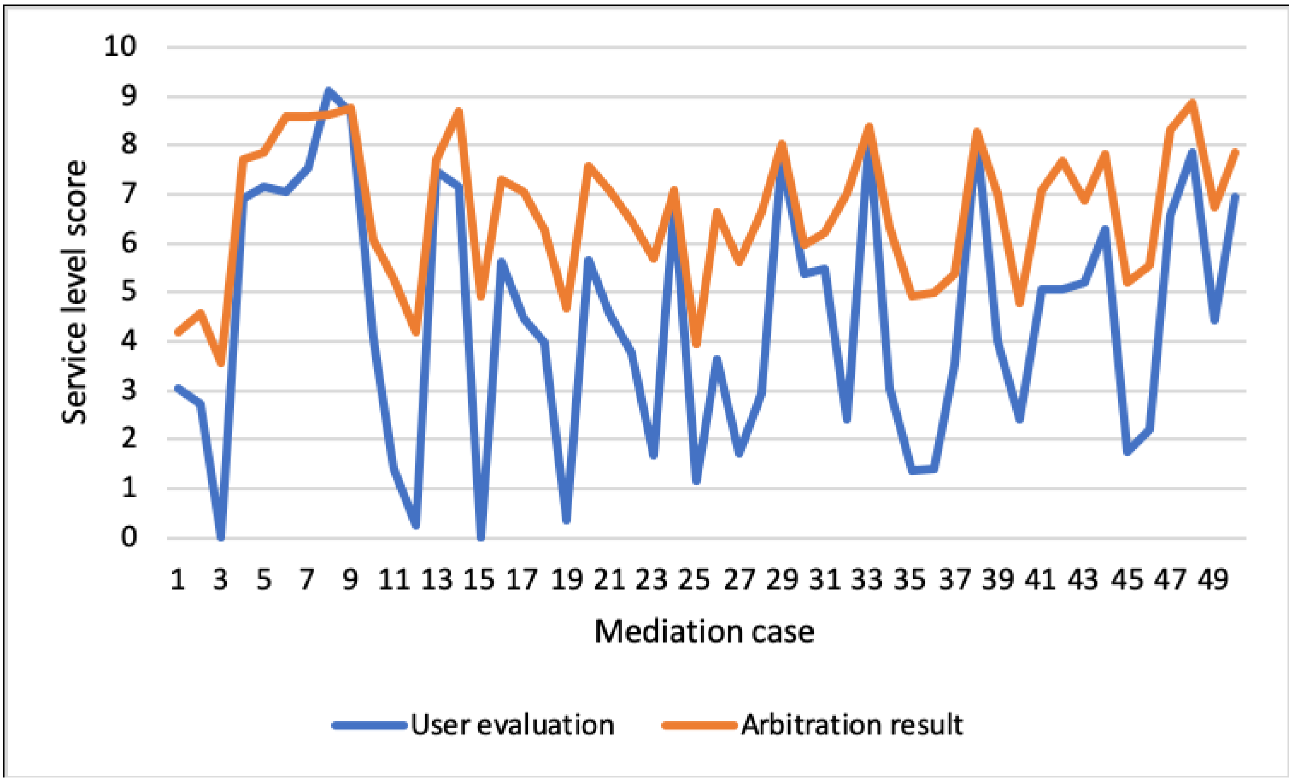
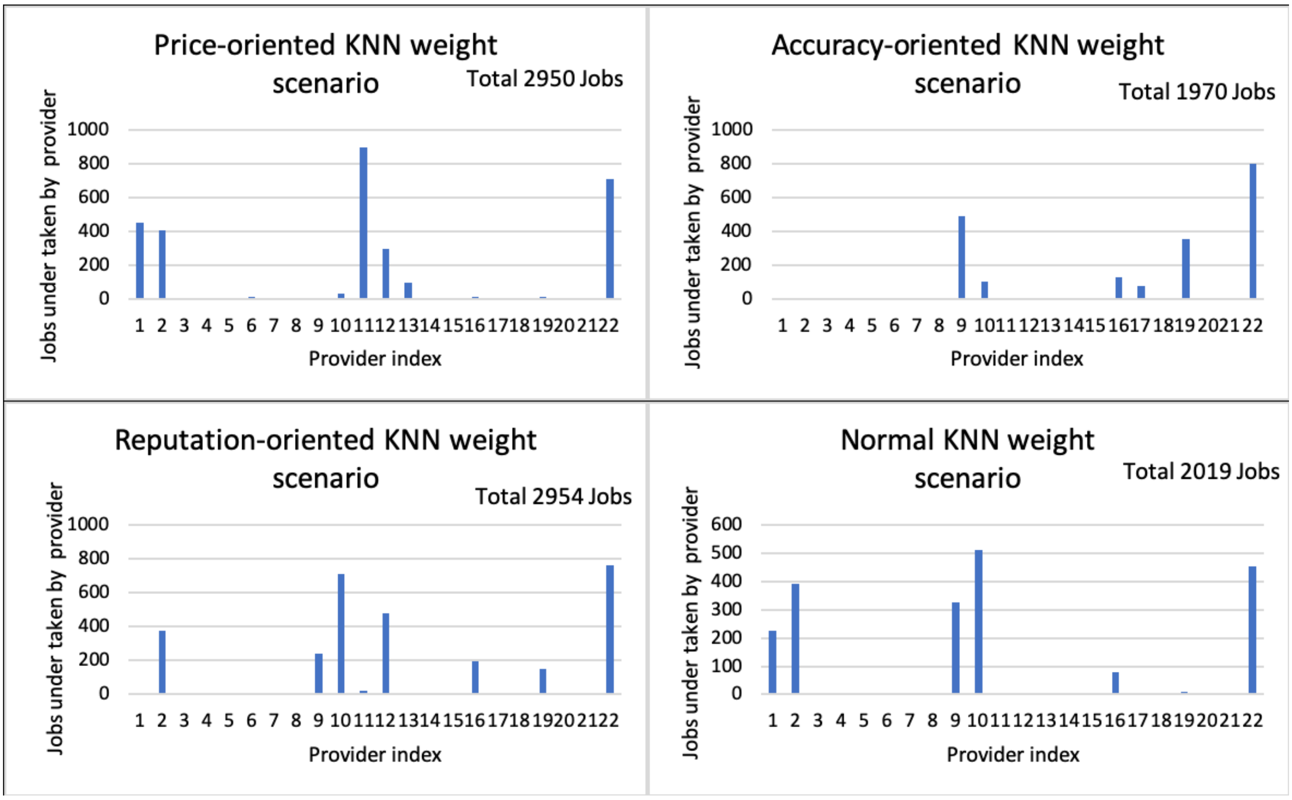


**Figure 14.** KNN weight sensitivity on provider selection. KNN: k-nearest neighbors.




[Figure 14](#) shows the number of service requests each provider fulfills under each KNN weight scenario. The diagram illustrates that provider 1 fulfills around 450 service requests under the price-oriented KNN weight scenario compared with 0 under the other two KNN weight scenarios. Under the price-oriented KNN weight scenario, Provider 11 fulfills approximately 900 jobs, which is the highest compared with other providers. Provider 22, who provides the most accurate service, fulfills the highest number of service requests under the accuracy-oriented and reputation-oriented KNN weight scenarios, while fulfilling the second highest number of service requests under the price-oriented and normal KNN weight scenarios. This indicates that provider 22 is preferred over all other providers overall. Provider 2 fulfills a significant number of service requests under the price-oriented, reputation-oriented, and normal KNN weight scenarios but fulfills a negligible number of requests under the accuracy-oriented KNN weight scenario. This is because provider 2 has advantages in price and service but is less satisfactory in accuracy. Providers 1 and 11 fulfill more service requests under the price-oriented KNN weight scenario than they fulfill under the other three scenarios. The implication is that these two providers offer a reasonable price for service, but they have poor performance in user service ratings and diagnostic accuracy. Providers 9 and 19 are not effective under the price-oriented KNN weight scenario, but they are effective in the other three KNN scenarios indicating they adopted an unsuccessful strategy in price competition. These observations reflect the nature of the system design: providers may have a disadvantage in one aspect, but they can still win the competition by offering attractive conditions in other aspects.

## Discussion

In this research, we propose a model of cloud health service sharing-based blockchain technology featuring resource sharing, consensus, global payment, and distributed ledger. This mechanism allows the proposed framework to have sufficient feasibility and be supported by an increasing number of participants. Based on the open source Ethereum blockchain system, we adopt a consortium or federated blockchain for the proposed framework. Solidity language is employed to develop smart contracts. A simulation study on breast cancer diagnosis is constructed. A recommendation algorithm is designed to find the proper providers for service requests. During the 500 hour simulated time of generated service requests, 9893 job requests are generated and fulfilled by 22 providers. All requests are fulfilled by service providers based on recommendations from the weighted KNN algorithm, and 9 providers take up the preponderance of service requests. User priority on KNN weights evidently affect system operation outcomes. Provider service evaluation scores stabilize as service requests increase during the simulation.

A service evaluation system is incorporated in which a novel arbitration mechanism is designed to address the issue of potential biased evaluations. Both the self-appraisal of the provider and the evaluation by the user are taken into account in the arbitration. This protects providers by mitigating negative evaluations of malicious users. Note that the arbitration process adopts a distributed decision model through voting to mediate conflict. Qualified arbitration committee members could be distributed worldwide. When arbitration is submitted, members of the arbitration committee are selected by the system randomly and the selected members constitute the arbitration committee only for the case. The system sends the summary information of the case to the committee members. The committee members vote based on the shreds of verified evidence. The system finally reaches a decision by a smart contract.

We feel that the proposed model has tremendous potential, and the current work represents only the first step by demonstrating its feasibility. Research extension is called for to better design and optimize the proposed model in the future. For instance, for the proposed model to be implemented in the real world, the design of a cryptocurrency system is of great importance. The issue has not been addressed in this study, and it is worth investigation. Also, with the introduction of blockchain technology in the proposed model, some security issues could be mitigated but new challenges could arise, such as the high overhead of blockchain technology, privacy of the transaction, majority attack, the scale of blockchain, current regulations issues, and the integrated cost problem [44]. The new challenges should be dealt with in the future. In terms of business models on such a proposed system, interesting topics include pricing strategies for providers under a variety of situations. In addition, new applications of the proposed model should be explored addressing medicine prescription tracking and health insurance claims.

## Authors' Contributions

XZ and JS contributed to the concept and design of the work and acquisition and analysis of the data. XZ contributed the code for the simulation. XZ and CL drafted the manuscript. JS revised the manuscript.

## Conflicts of Interest

None declared.

## References

1. World Health Organization. 2018. Public spending on health: a closer look at global trends URL: https://www.who.int/health_financing/documents/health-expenditure-report-2018/en/ [accessed 2019-06-22]
2. Nie L, Li T, Akbari M, Shen J, Chua TS. Comprehensive vertical search for healthcare domain. 2014 Presented at: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval; 2014; Gold Coast p. 1245-1246. [doi: 10.1145/2600428.2611176]
3. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 2017;7(4):351-367 [FREE Full text] [doi: 10.1007/s12553-017-0179-1] [Medline: 29308344]
4. Wong MC, Yee KC, Nøhr C. Socio-technical considerations for the use of blockchain technology in healthcare. Stud Health Technol Inform 2018;247:636-640. [Medline: 29678038]
5. Adler-Milstein J, Bates DW. Paperless healthcare: progress and challenges of an IT-enabled healthcare system. Business Horizons 2010 Mar;53(2):119-130. [doi: 10.1016/j.bushor.2009.10.004]
6. Zhang P, Walker M, White J, Schmidt D, Lenz G. Metrics for assessing blockchain-based healthcare decentralized apps. 2017 Presented at: e-Health Networking, Applications and Services (Healthcom), IEEE 19th International Conference; 2017; Dalian p. 1-4. [doi: 10.1109/HealthCom.2017.8210842]
7. Zhang P, Schmidt D, White J, Lenz G. Blockchain technology use cases in healthcare. In: Raj P, Ganesh CD, editors. Advances in Computers. Blockchain Technology: Platforms, Tools and Use Cases, Volume 111. Cambridge: Academic Press; 2018.
8. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. 2019. URL: https://bitcoin.org/bitcoin.pdf [accessed 2019-06-22]
9. Rifi N, Rachkidi E, Agoulmine N, Taher N. Towards using blockchain technology for eHealth data access management. In: Advances in Biomedical Engineering. 2017 Presented at: ICABME, Fourth International Conference; 2017; Beirut p. 1-4. [doi: 10.1109/ICABME.2017.8167555]
10. Xia Q, Sifah E, Smahi A, Amofa S, Zhang X. BBDS: blockchain-based data sharing for electronic medical records in cloud environments. Information 2017 Apr 17;8(2):44. [doi: 10.3390/info8020044]
11. Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. 2017 Presented at: Personal, Indoor, Mobile Radio Communications (PIMRC), IEEE 28th Annual International Symposium; 2017; Montreal p. 1-5. [doi: 10.1109/PIMRC.2017.8292361]
12. Alexaki S, Alexandris G, Katos V, Petroulakis E. Blockchain-based electronic patient records for regulated circular healthcare jurisdictions. 2018 Presented at: IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD); 2018; Barcelona p. 1-6. [doi: 10.1109/CAMAD.2018.8514954]
13. Ji Y, Zhang J, Ma J, Yang C, Yao X. BMPLS: blockchain-based multi-level privacy-preserving location sharing scheme for telecare medical information systems. J Med Syst 2018 Jun 30;42(8):147. [doi: 10.1007/s10916-018-0998-2] [Medline: 29961160]
14. Dey T, Jaiswal S, Sunderkrishnan S, Katre N. A medical use case of Internet of Things and blockchain. 2017 Presented at: International Conference on Intelligent Sustainable Systems (ICISS); 2017; Palladam p. 486-491. [doi: 10.1109/ISS1.2017.8389459]

15. Li H, Zhu L, Shen M, Gao F, Tao X, Liu S. Blockchain-based data preservation system for medical data. J Med Syst 2018 Jun 28;42(8):141. [doi: 10.1007/s10916-018-0997-3] [Medline: 29956058]
16. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustainable Cities Soc 2018 May;39:283-297. [doi: 10.1016/j.scs.2018.02.014]
17. 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]
18. 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]
19. 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. [doi: 10.1109/ACCESS.2018.2801266]
20. Zhang P, White J, Schmidt D, Lenz G. Design of blockchain-based apps using familiar software patterns with a healthcare focus. 2017 Presented at: Proceedings of the 24th Conference on Pattern Languages of Programs. The Hillside Group; 2017; Vancouver p. 23-25.
21. Brogan J, Baskaran I, Ramachandran N. Authenticating health activity data using distributed ledger technologies. Comput Struct Biotechnol J 2018;16:257-266. [doi: 10.1016/j.csbj.2018.06.004] [Medline: 30101004]
22. Tian H, He J, Ding Y. Medical data management on blockchain with privacy. J Med Syst 2019;43(2):26. [Medline: 30603816]
23. Mettler M. Blockchain technology in healthcare: The revolution starts here. 2016 Presented at: 2016 IEEE 18th International Conference; 2016; Munich p. 1-3. [doi: 10.1109/HealthCom.2016.7749510]
24. Dhillon V, Metcalf D, Hooper M. Blockchain in science. In: Blockchain Enabled Applications: Understand the Blockchain Ecosystem and How to Make it Work for You. Berkeley: Apress; 2017:111-124.
25. Taylor B, Hare G. Extended blockchains for event tracking and management. 2018. URL: https://patents.justia.com/patent/20180130050 [accessed 2019-06-22]
26. Shae Z, Tsai J. On the design of a blockchain platform for clinical trial and precision medicine. 2017 Presented at: IEEE 37th International Conference on Distributed Computing Systems (ICDCS); 2017; Atlanta p. 1972-1980. [doi: 10.1109/ICDCS.2017.61]
27. Suzuki S, Murai J. Blockchain as an audit-able communication channel. 2017 Presented at: IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); 2017; Turin p. 516-522. [doi: 10.1109/COMPSAC.2017.72]
28. Tsai J. Transform blockchain into distributed parallel computing architecture for precision medicine. 2018 Presented at: IEEE 38th International Conference on Distributed Computing Systems (ICDCS); 2018; Vienna p. 1290-1299. [doi: 10.1109/ICDCS.2018.00129]
29. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: transparency and traceability of consent. F1000Res 2017;6:66 [FREE Full text] [doi: 10.12688/f1000research.10531.5] [Medline: 29167732]
30. Maslova N. Blockchain: disruption and opportunity. Strategic Finance 2018;100(1):24-30 [FREE Full text]
31. Adler J, Berryhill R, Veneris A, Poulos Z, Veira N, Kastania A. Astraea: a decentralized blockchain oracle. 2018 Presented at: IEEE International Conference on Internet of Things (iThings), IEEE Green Computing and Communications (GreenCom), IEEE Cyber, Physical and Social Computing (CPSCom), IEEE Smart Data (SmartData); 2018; Halifax. [doi: 10.1109/Cybermatics_2018.2018.00207]
32. De Angelis A, Baldoni R, Lombardi F, Margheri A. Pbft vs proof-of-authority: applying the cap theorem to permissioned blockchain. 2018 Presented at: Proceedings of the Second Italian Conference on Cyber Security; 2018; Milan p. 6-9.
33. Mangasarian OL, Street WN, Wolberg WH. Breast cancer diagnosis and prognosis via linear programming. Operations Res 1995 Aug;43(4):570-577. [doi: 10.1287/opre.43.4.570]
34. Wolberg W, Street W, Heisey D. Computer-derived nuclear features distinguish malignant from benign breast cytology. Hum Pathol 1995;26(7):792. [Medline: 7628853]
35. Bamakan SMH, Gholami P. A novel feature selection method based on an integrated data envelopment analysis and entropy model. Procedia Comput Sci 2014;31:632-638. [doi: 10.1016/j.procs.2014.05.310]
36. Xue B, Zhang M, Browne W. New fitness functions in binary particle swarm optimisation for feature selection. 2012 Presented at: Evolutionary Computation (CEC); 2012; Brisbane p. 1-8. [doi: 10.1109/CEC.2012.6256617]
37. Maldonado S, Weber R, Basak J. Simultaneous feature selection and classification using kernel-penalized support vector machines. Inf Sci 2011 Jan;181(1):115-128. [doi: 10.1016/j.ins.2010.08.047]
38. Miao D, Gao C, Zhang N, Zhang Z. Diverse reduct subspaces based co-training for partially labeled data. Int J Approx Reasoning 2011 Nov;52(8):1103-1117. [doi: 10.1016/j.ijar.2011.05.006]
39. Luukka P, Leppälampi T. Similarity classifier with generalized mean applied to medical data. Comput Biol Med 2006 Sep;36(9):1026-1040. [doi: 10.1016/j.compbiomed.2005.05.008]
40. Sheikhpour R, Sarram MA, Sheikhpour R. Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer. Applied Soft Computing 2016 Mar;40:113-131. [doi: 10.1016/j.asoc.2015.10.005]

41. Kaur T, Saini BS, Gupta S. A novel feature selection method for brain tumor MR image classification based on the Fisher criterion and parameter-free Bat optimization. Neural Comput & Applic 2017 Feb 21;29(8):193-206. [doi: 10.1007/s00521-017-2869-z]
42. Samworth RJ. Optimal weighted nearest neighbour classifiers. Ann Statist 2012 Oct;40(5):2733-2763. [doi: 10.1214/12-AOS1049]
43. Peterson RA, Wilson WR. Measuring customer satisfaction: fact and artifact. JAMS 1992 Dec;20(1):61-71. [doi: 10.1007/BF02723476]
44. Lin I, Liao T. A survey of blockchain security issues and challenges. IJ Network Security 2017;19(5):653. [doi: 10.3934/mfc.2018007]

## Abbreviations

**AI:** artificial intelligence
**CHC:** cloud health coin
**DApp:** decentralized application
**EHR:** electronic health record
**KNN:** k-nearest neighbors
**PoA:** Proof of Authority

*Edited by P Zhang, K Clauson; submitted 21.02.19; peer-reviewed by C Reis, JT te Gussinklo; comments to author 27.04.19; revised version received 23.05.19; accepted 20.06.19; published 23.07.19.*

*Please cite as:*
*Zhu X, Shi J, Lu C*
*Cloud Health Resource Sharing Based on Consensus-Oriented Blockchain Technology: Case Study on a Breast Tumor Diagnosis Service*
*J Med Internet Res 2019;21(7):e13767*
*URL: http://www.jmir.org/2019/7/e13767/*
*doi:10.2196/13767*
*PMID:31339106*

Viewpoint

# Reducing Patient Loneliness With Artificial Agents: Design Insights From Evolutionary Neuropsychiatry

Kate Loveys[1], MSc; Gregory Fricchione[2,3], MD; Kavitha Kolappa[2], MD; Mark Sagar[4], PhD; Elizabeth Broadbent[1], PhD

[1]Department of Psychological Medicine, The University of Auckland, Auckland, New Zealand

[2]Benson-Henry Institute for Mind Body Medicine, Massachusetts General Hospital, Boston, MA, United States

[3]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

[4]Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand

**Corresponding Author:**
Kate Loveys, MSc
Department of Psychological Medicine
The University of Auckland
Auckland City Hospital, Level 12 Support Building
85 Park Road, Grafton
Auckland, 1023
New Zealand
Phone: 64 9 373 7599 ext 84340
Email: k.loveys@auckland.ac.nz

## Abstract

Loneliness is a growing public health issue that substantially increases the risk of morbidity and mortality. Artificial agents, such as robots, embodied conversational agents, and chatbots, present an innovation in care delivery and have been shown to reduce patient loneliness by providing social support. However, similar to doctor and patient relationships, the quality of a patient's relationship with an artificial agent can impact support effectiveness as well as care engagement. Incorporating mammalian attachment-building behavior in neural network processing as part of an agent's capabilities may improve relationship quality and engagement between patients and artificial agents. We encourage developers of artificial agents intended to relieve patient loneliness to incorporate design insights from evolutionary neuropsychiatry.

***(J Med Internet Res 2019;21(7):e13664)*** doi:10.2196/13664

**KEYWORDS**

loneliness; neuropsychiatry; biological evolution; psychological bonding; interpersonal relations; artificial intelligence; social support; eHealth

## Introduction

Artificial agents, such as robots and chatbots, are currently being developed to provide companionship and assist patients with health care needs [1,2]. The purpose of this paper is to argue that incorporating mammalian attachment-building behavior into agent design may increase agent effectiveness. First, the paper presents loneliness as a growing public health issue and discusses the promise of social support interventions for treating loneliness. The paper then describes recent social support interventions delivered by artificial agents. Finally, the paper presents insights from evolutionary neuropsychiatry and describes mammalian attachment-building behaviors that may be included in artificial agents to promote patient engagement.

## Loneliness, Social Connection, and Health

Loneliness is a widespread global health issue that approximately affects a third of people in industrialized countries [3]. Loneliness refers to a subjective state of social isolation in which the individual perceives a mismatch between ideal and actual social relations [4]. According to a recent report by the Jo Cox Commission, almost a quarter of parents with young children felt *always or often lonely*, more than a third of those aged over 75 years reported feelings of loneliness out of their control, and, in one year, more than 4000 children reported feeling *unbearably lonely* in the United Kingdom alone [5]. Hospitalized patients are at a particularly high risk of loneliness [6].

Although the occasional feeling of loneliness acts as an adaptive signal to seek social interaction, chronic loneliness can be detrimental to health. Loneliness increases mortality risk by 32% [7], a degree comparable with smoking 15 cigarettes daily [8]. Loneliness elevates the risk of many morbidities, including stress-related conditions (coronary heart disease, stroke, and high blood pressure) [9], and psychiatric illnesses (major depression, psychosis, and suicide) [10]. Loneliness places a significant burden on health care systems through increased health care utilization [11], costing an estimated additional US $6.7 billion per year for older adults alone [12].

Social connection refers to quality social relations characterized by perceived belongingness and closeness [13]. Greater social connection is needed to counteract the significant detrimental effects of loneliness on health and longevity and relieve the burden on health care systems.

## Social Support Interventions

Many researchers are testing social support interventions as a means to improve social connection and reduce loneliness. Social support refers to a functional exchange of emotional, informational, or practical aid between individuals [14]. One strategy to provide social support is via community-level interventions. An example is the Campaign to End Loneliness, which promotes small actions of social connection between strangers in the United Kingdom. Other examples include the Reconnections Service (which links older adults to social activities), computer skills training for older adults to engage with others on the Web [15], peer support [14], and altering the environment to be more conducive to social interaction by providing pedestrian-focused public spaces [16].

One way that social support interventions protect health is by buffering against the impact of stress on the body [17]. Chronic stress increases inflammation [18], lowers heart rate variability [19], and impairs immune response [20], which increase the risk of physical and mental morbidities [18,20,21]. Social support can reduce sympathetic nervous system activation [22], increase oxytocin secretion, and suppress cortisol release [23], which reduce the impact of chronic stress [17]. Social support can also indirectly benefit health through the provision of health information, treatment adherence encouragement, or practical support [24].

Interventions to improve social support have generally been shown to have benefits for health and well-being, including reduced stress [23], lower anxiety and depression [25], decreased alcohol consumption [26], and improvements in wound healing [27], treatment adherence [28], myocardial infarction recurrences, and mortality [29]. A systematic review of 100 studies concluded that social support interventions generally provide health benefits irrespective of the type of support provided in the intervention and whether interventions were delivered to an individual or group or were professionally led or provided by peers [14]. However, the effectiveness of social support may be impacted by components of relationship quality between the patient and support provider, such as social closeness. Social closeness refers to a relationship quality where partners regularly engage in intimate behaviors such as support, self-disclosure, and shared activities [30]. Characteristics of the partner can also affect how close a connection is formed; these include perceived familiarity [31], warmth [30], and empathic accuracy [32].

## Social Support From Artificial Agents

Traditional social support interventions may not always be available or desirable. In some situations, faced with the absence of human connection, artificial agents may provide support akin to human social support to benefit health. Artificial agents in health care may act as transitional objects that help patients to cope with feelings of loneliness and the depressive anxiety that often accompanies severe illness and end of life experiences [33,34]. Artificial agents have been shown to reduce feelings of loneliness [35], as well increase interrelatedness, either through direct interactions with the agent or by triggering conversations between humans that might not have otherwise occurred [1].

In addition to social benefits, artificial agents have been shown to exert positive effects on physical and mental health. Paro (Intelligent System Co, Ltd, Japan), a companion robot in the form of a fluffy baby harp seal, has been shown to improve mood [36] and reduce depression symptoms for people with dementia [37]. Paro was designed with big eyes and soft fur to encourage users to feel affectionate toward it like a real baby animal. iRobi (Yujin Robot Co, Ltd, Korea), a robotic homecare companion, significantly improved medication adherence and rehabilitation exercise frequency for patients with chronic obstructive pulmonary disease through providing information and reminders [38]. Conversational agents have demonstrated benefits for mental health, such as reduced depression and anxiety symptoms [2], and other forms of artificial companionship are being explored in the context of health, including Alexa (a virtual voice assistant made by Amazon) [39].

The characteristics of an artificial agent providing social support may affect the success of an intervention. If an agent closely models realistic human interactions, this may increase patients' willingness to develop social closeness with the agent [40]. A very high degree of human likeness improves perceptions of agents' social characteristics [41], and appropriate use of human-like verbal and nonverbal relational cues improve an agent's relationship with users [42]. The most important behaviors for developing social closeness and support between humans and artificial agents remain to be determined. Behaviors from the natural world may provide some promising design strategies.

## Design Insights From Evolutionary Neuropsychiatry

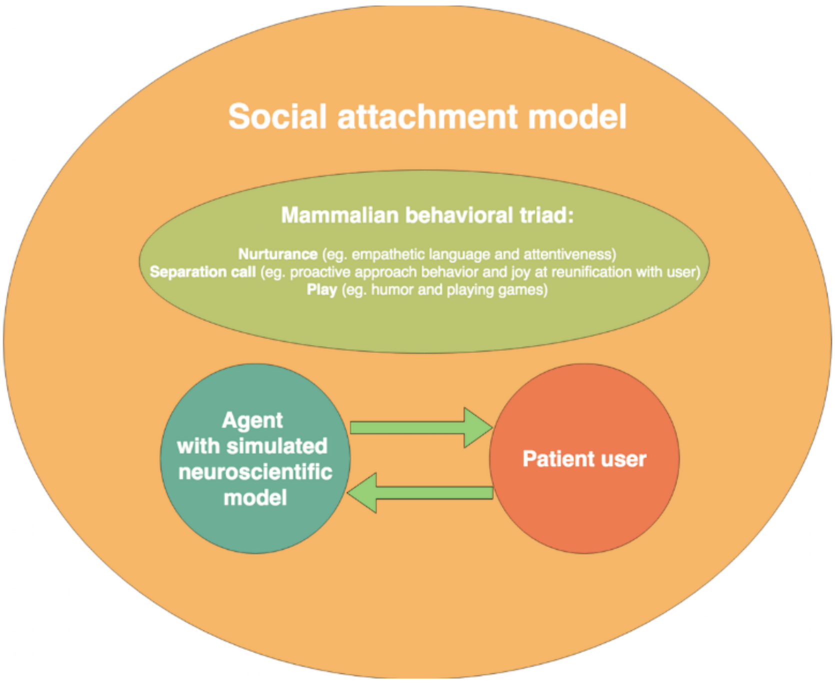
We propose that evolutionary neuropsychiatry offers important insights for the design of artificial agents to provide complementary support meant to be additive and not substitutive to human support. Aspects of mammalian brain evolution that enabled social attachments provide direction to engineers as to the necessary internal structure, processes, and output required of systems to appropriately elicit attachment in a way that maximally supports human users.

MacLean studied aspects of brain evolution across reptiles and mammals and found that for mammalian brain evolution, specifically, particular structures evolved that enabled the mammalian behavioral triad [43]. The mammalian behavioral triad comprises maternal nurturance, the separation call, and social play. These behaviors serve the purpose of strengthening social attachment, which is the mammalian survival strategy.

For the mammalian behavioral triad to be possible, as well as the ability for mammals to select attachment or separation as a response to environmental objects, certain brain structures and loops evolved. The *protolimbic* loop evolved to manage attachment to food and reproductive objects [33]. Two primordial moieties assist the protolimbic loop: the hippocampocentric moiety specifies where the organism is located in relation to objects in the environment, whereas the olfactocentric moiety classifies objects that are located within the environment [33]. The mammalian brain evolved in such a way that these 2 moieties converged in paralimbic cortical zones, namely, the anterior cingulate cortex (ACC), the medial orbitofrontal cortex (mOFC), and the anterior insula. This created a terminal zone for a *paralimbic* basal ganglia thalamocortical circuit. In the paralimbic loop, the ACC works with the mOFC to synthesize and make emotional and cognitive classifications of input to inform decisions about whether to separate from or attach to an object. The convergence of these areas created a response selecting area which can be traced to the primary separation challenge attachment solution paradigm [33]. This paradigm indicates requirements for mammals to attach not only to sources of metabolic energy and reproductive success but also to sources of social support for survival. The evolution of these particular structures and their convergence in the mammalian brain provided the basis for social attachment, both within and between mammalian species.

We propose that artificial agents designed for social support provision be created with internal models of the neural structures, processes, and output that evolved to enable genuine social attachment between mammalian species. This involves creating agents with environmental sensors, classifiers for incoming data on emotion and attachment behavior, and interaction memory with a user, along with the behavioral capacity to produce the mammalian behavioral triad. For example, nurturance could be shown by attentiveness and use of empathetic language; the separation call could be shown through the proactive arrangement of another meeting in the future; and play could be demonstrated through the use of humor. A model of social attachment between artificial agents and patients is shown (Figure 1). The simulation of biological processes necessary for producing such behaviors in artificial agents is complex, but substantial progress has been made toward linking neuroscience models with computer graphic interfaces for creating life-like facial expressions during interactions [40].

**Figure 1.** Model of behavior that may increase patient engagement with artificial agents according to evolutionary neuropsychiatry.




Ideally, patients and agents should form a reciprocal attachment over time with repeated interactions. It is an aspiration that future artificial agents in health care may have embedded the capacity to produce an efficacious facsimile of social attachment, which may enhance the potency of an agent's social support, reduce patient loneliness, and improve patient engagement with care.

Although we advocate that design of artificial agents be inspired by the evolutionary neuropsychiatry of social attachment, we express the caveat that these agents would serve only as adjuvant social support boosters and would not be designed as substitutive for genuine human attachments. We also acknowledge that the design of artificial agents intended for a high degree of human interaction is a complex, sensitive issue that requires multidisciplinary discussion by diverse stakeholders and demographic groups, particularly in relation to ethics and evaluation [44,45]. Further consideration of safeguards embedded into agent design and implementation, as well as ongoing evaluations using validated metrics, is necessary to ensure social connection with artificial agents is beneficial for patients.

## Conflicts of Interest

MS is the CEO of Soul Machines (an artificial intelligence company), which supports KL with a PhD stipend, and contracts EB for consultancy work.

## References

1. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc 2013 Sep;14(9):661-667. [doi: 10.1016/j.jamda.2013.02.007] [Medline: 23545466]
2. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
3. Cacioppo JT, Cacioppo S. The growing problem of loneliness. Lancet 2018 Dec 03;391(10119):426 [FREE Full text] [doi: 10.1016/S0140-6736(18)30142-9] [Medline: 29407030]
4. Cacioppo S, Grippo AJ, London S, Goossens L, Cacioppo JT. Loneliness: clinical import and interventions. Perspect Psychol Sci 2015 Mar;10(2):238-249 [FREE Full text] [doi: 10.1177/1745691615570616] [Medline: 25866548]
5. Jo Cox Commission. 2017. Combatting loneliness one conversation at a time URL: https://www.jocoxloneliness.org/pdf/a_call_to_action.pdf [WebCite Cache ID 75PS537HU]
6. Karhe L, Kaunonen M. Patient experiences of loneliness: an evolutionary concept analysis. ANS Adv Nurs Sci 2015;38(4):E21-E34. [doi: 10.1097/ANS.0000000000000096] [Medline: 26517346]
7. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
8. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010 Jul 27;7(7):e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
9. Valtorta NK, Kanaan M, Gilbody S, Ronzi S, Hanratty B. Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. Heart 2016 Dec 01;102(13):1009-1016 [FREE Full text] [doi: 10.1136/heartjnl-2015-308790] [Medline: 27091846]
10. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
11. Gerst-Emerson K, Jayawardhana J. Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults. Am J Public Health 2015 May;105(5):1013-1019. [doi: 10.2105/AJPH.2014.302427] [Medline: 25790413]
12. Shaw JG, Farid M, Noel-Miller C, Joseph N, Houser A, Asch SM, et al. Social isolation and medicare spending: among older adults, objective social isolation increases expenditures while loneliness does not. J Aging Health 2017 Dec;29(7):1119-1143 [FREE Full text] [doi: 10.1177/0898264317703559] [Medline: 29545676]
13. Lee RM, Draper M, Lee S. Social connectedness, dysfunctional interpersonal behaviors, and psychological distress: testing a mediator model. J Couns Psychol 2001;48(3):310-318. [doi: 10.1037/0022-0167.48.3.310]
14. Hogan BE, Linden W, Najarian B. Social support interventions: do they work? Clin Psychol Rev 2002 Apr;22(3):383-442. [Medline: 17201192]
15. Choi M, Kong S, Jung D. Computer and internet interventions for loneliness and depression in older adults: a meta-analysis. Healthc Inform Res 2012 Sep;18(3):191-198 [FREE Full text] [doi: 10.4258/hir.2012.18.3.191] [Medline: 23115742]
16. Leyden KM. Social capital and the built environment: the importance of walkable neighborhoods. Am J Public Health 2003 Sep;93(9):1546-1551. [doi: 10.2105/ajph.93.9.1546] [Medline: 12948978]
17. Lepore SJ. Problems and prospects for the social support-reactivity hypothesis. Ann Behav Med 1998;20(4):257-269. [doi: 10.1007/BF02886375] [Medline: 10234419]
18. Cohen S, Janicki-Deverts D, Doyle WJ, Miller GE, Frank E, Rabin BS, et al. Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk. Proc Natl Acad Sci U S A 2012 Apr 17;109(16):5995-5999 [FREE Full text] [doi: 10.1073/pnas.1118355109] [Medline: 22474371]
19. Kim H, Cheon E, Bai D, Lee YH, Koo B. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
20. Padgett DA, Glaser R. How stress influences the immune response. Trends Immunol 2003 Aug;24(8):444-448. [Medline: 12909458]

21. McEwen BS. Protection and damage from acute and chronic stress: allostasis and allostatic overload and relevance to the pathophysiology of psychiatric disorders. Ann N Y Acad Sci 2004 Dec;1032:1-7. [doi: 10.1196/annals.1314.001] [Medline: 15677391]
22. Kiecolt-Glaser JK, Gouin J, Hantsoo L. Close relationships, inflammation, and health. Neurosci Biobehav Rev 2010 Sep;35(1):33-38 [FREE Full text] [doi: 10.1016/j.neubiorev.2009.09.003] [Medline: 19751761]
23. Heinrichs M, Baumgartner T, Kirschbaum C, Ehlert U. Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. Biol Psychiatry 2003 Dec 15;54(12):1389-1398. [Medline: 14675803]
24. House J, Kahn R. Measures and concepts of social support. In: Social Support and Health. New York: Academic Press; 1985.
25. Thompson DR. A randomized controlled trial of in-hospital nursing support for first time myocardial infarction patients and their partners: effects on anxiety and depression. J Adv Nurs 1989 Apr;14(4):291-297. [doi: 10.1111/j.1365-2648.1989.tb03416.x]
26. Irvine MJ, Logan AG. Relaxation behavior therapy as sole treatment for mild hypertension. Psychosom Med 1991;53(6):587-597. [Medline: 1758944]
27. Robinson H, Ravikulan A, Nater UM, Skoluda N, Jarrett P, Broadbent E. The role of social closeness during tape stripping to facilitate skin barrier recovery: preliminary findings. Health Psychol 2017;36(7):619-629. [doi: 10.1037/hea0000492]
28. DiMatteo MR. Social support and patient adherence to medical treatment: a meta-analysis. Health Psychol 2004 Mar;23(2):207-218. [doi: 10.1037/0278-6133.23.2.207] [Medline: 15008666]
29. Frasure-Smith N, Prince R. The ischemic heart disease life stress monitoring program: impact on mortality. Psychosom Med 1985;47(5):431-445. [Medline: 4059478]
30. Parks MR, Floyd K. Meanings for closeness and intimacy in friendship. J Soc Pers Relat 2016 Jun 30;13(1):85-107. [doi: 10.1177/0265407596131005]
31. Robles TF. Stress, social support, and delayed skin barrier recovery. Psychosom Med 2007 Nov;69(8):807-815. [doi: 10.1097/PSY.0b013e318157b12e] [Medline: 17942836]
32. Kraus MW. Voice-only communication enhances empathic accuracy. American Psychologist 2017 Oct;72(7):644-654. [doi: 10.1037/amp0000147]
33. Fricchione GL. Compassion and Healing in Medicine and Society: On the Nature and Use of Attachment Solutions to Separation Challenges. Baltimore, MD: Johns Hopkins University Press; 2011.
34. Winnicott DW. Transitional objects and transitional phenomena; a study of the first not-me possession. Int J Psychoanal 1953;34(2):89-97. [Medline: 13061115]
35. Ring L, Shi L, Totzke K, Bickmore T. Social support agents for older adults: longitudinal affective computing in the home. J Multimodal User Interfaces 2014 Jun 18;9(1):79-88. [doi: 10.1007/s12193-014-0157-0]
36. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. Use of a therapeutic, socially assistive pet robot (PARO) in improving mood and stimulating social interaction and communication for people with dementia: study protocol for a randomized controlled trial. JMIR Res Protoc 2015 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]
37. Jøranson N, Pedersen I, Rokstad AM, Ihlebæk C. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. J Am Med Dir Assoc 2015 Oct;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002]
38. Broadbent E, Garrett J, Jepsen N, Li OV, Ahn HS, Robinson H, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
39. Chung AE, Griffin AC, Selezneva D, Gotz D. Health and fitness apps for hands-free voice-activated assistants: content analysis. JMIR Mhealth Uhealth 2018 Sep 24;6(9):e174 [FREE Full text] [doi: 10.2196/mhealth.9705] [Medline: 30249581]
40. Sagar M, Seymour M, Henderson A. Creating connection with autonomous facial animation. Commun ACM 2016 Dec 01;59(12):82-91. [doi: 10.1145/2950041]
41. Broadbent E, Kumar V, Li X, Sollers J, Stafford RQ, MacDonald BA, et al. Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. PLoS ONE 2013 Aug 28;8(8):e72589. [doi: 10.1371/journal.pone.0072589]
42. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 2005 Jun 01;12(2):293-327. [doi: 10.1145/1067860.1067867]
43. MacLean PD. Brain evolution relating to family, play, and the separation call. Arch Gen Psychiatry 1985 Apr 01;42(4):405. [doi: 10.1001/archpsyc.1985.01790270095011]
44. de Graaf MM. An ethical evaluation of human-robot relationships. Int J Soc Robot 2016;8(4):589-598. [doi: 10.1007/s12368-016-0368-5]
45. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_brochure_v2.pdf [WebCite Cache ID 75PRxfM5p]

## Abbreviations

**ACC:** anterior cingulate cortex
**mOFC:** medial orbitofrontal cortex

*Edited by C Eichenberg; submitted 17.02.19; peer-reviewed by S Ge, A Scoglio; comments to author 24.03.19; revised version received 07.04.19; accepted 08.04.19; published 08.07.19.*

*Please cite as:*
*Loveys K, Fricchione G, Kolappa K, Sagar M, Broadbent E*
*Reducing Patient Loneliness With Artificial Agents: Design Insights From Evolutionary Neuropsychiatry*
*J Med Internet Res 2019;21(7):e13664*
*URL: https://www.jmir.org/2019/7/e13664/*
*doi:10.2196/13664*
*PMID:31287067*

Review

# Use of Social Robots in Mental Health and Well-Being Research: Systematic Review

Arielle AJ Scoglio[1], BA, MPH; Erin D Reilly[1], MEd, PhD; Jay A Gorman[1], PhD; Charles E Drebing[1], PhD

Social & Community Reintegration Research Program, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

**Corresponding Author:**
Arielle AJ Scoglio, BA, MPH
Social & Community Reintegration Research Program
Edith Nourse Rogers Memorial Veterans Hospital
200 Springs Road, Mailstop 116A
Bedford, MA,
United States
Phone: 1 781 687 4527
Email: arielle.scoglio@va.gov

## Abstract

**Background:** Technology-assisted clinical interventions are increasingly common in the health care field, often with the proposed aim to improve access to and cost-effectiveness of care. Current technology platforms delivering interventions are largely mobile apps and online websites, although efforts have been made to create more personalized and embodied technology experiences. To extend and improve on these platforms, the field of robotics has been increasingly included in conversations of how to deliver technology-assisted, interactive, and responsive mental health and psychological well-being interventions. Socially assistive robots (SARs) are robotic technology platforms with audio, visual, and movement capabilities that are being developed to interact with individuals socially while also assisting them with management of their physical and psychological well-being. However, little is known about the empirical evidence or utility of using SARs in mental health interventions.

**Objective:** The review synthesizes and describes the nascent empirical literature of SARs in mental health research and identifies strengths, weaknesses, and opportunities for improvement in future research and practice.

**Methods:** Searches in Medline, PsycINFO, PsycARTICLES, PubMed, and IEEE Xplore yielded 12 studies included in the final review after applying inclusion and exclusion criteria. Abstract and full-text reviews were conducted by two authors independently.

**Results:** This systematic review of the literature found 5 distinct SARs used in research to investigate the potential for this technology to address mental health and psychological well-being outcomes. Research on mental health applications of SARs focuses largely on elderly dementia patients and relies on usability pilot data with methodological limitations.

**Conclusions:** The current SARs research in mental health use is limited in generalizability, scope, and measurement of psychological outcomes. Opportunities for expansion of research in this area include diversifying populations studied, SARs used, clinical applications, measures used, and settings for those applications.

***(J Med Internet Res 2019;21(7):e13322)*** doi:10.2196/13322

**KEYWORDS**

social robotics; socially assistive robots; mental health; interventions

## Introduction

### Overview

There is a well-documented gap between individuals in need of support for mental health concerns and those who receive care [1-4]. To address the treatment disparity between individuals in need of psychiatric care and those who receive it, the field of mental health has expanded from offering exclusively in-office clinical care, including interventions such as psychotherapy, care management, and medication, to offering treatment in a wide array of settings and with varied interface platforms. The use of technology in providing at-home care options have identified both potential [5] and challenges [6].

One study found that technology-assisted support and treatments may be appealing to potential patients and health centers because telehealth delivery methods can improve access and are often cost effective [7]. There are multiple reasons why individuals may not be able to access needed care, including lack of available therapists [8], lack of transportation [9], stigma around engaging in mental health care [10], and financial barriers [11]. Mobile intervention options can also be used to extend in-person clinical treatments such as cognitive-behavioral therapy for insomnia or posttraumatic stress disorder treatments to the home [12,13]. Similarly, online interventions have the potential to reach patients who might otherwise not have access to mental health or other clinical care, as 82% of adults in the United States have access to either the internet at home or wireless mobile devices [14]. At-home technology platforms have attempted to meet the needs of such persons, to either replace or supplement in-person treatments [5,15].

Following initial program development, developers often purport that technology-supported behavioral interventions can be more easily implemented than in-person options, without the limitations of travel, local resources, training new practitioners to implement a treatment model, or monitoring treatment fidelity [16]. However, creating and implementing such systems can be difficult. For instance, real-world use of some mHealth tools by clinics and consumers remains low [17] even after the tools have been assessed for usability issues. At the same time, research has shown that mHealth tools may help clients self-manage their own treatment and goals across multiple diagnoses related to chronic physical and mental health challenges using apps, clinical portals, and texting interfaces [18,19]. Some research suggests that the utilization gap between developed technologies and their intended consumers may be related to engagement [20].

Although initial research appears promising, mHealth technology has documented limitations regarding treatment engagement [21,22]. Current mHealth technologies often rely on intervention strategies with minimal personalization and interactions, including mobile phone apps and one-way texting, which lack key factors of mental health interventions: real-time interactive engagement, simple user experiences, transdiagnostic capabilities within one platform, and personalized feedback [23]. Only recently has there been a push to develop and assess how socially interactive technologies, such as computer-animated virtual therapists, can be leveraged for mHealth interventions to support accountability, provide continuously tailored feedback, and form a social relationship to successfully impact client wellness [24]. Consequently, there is both a need for and room to improve the mHealth platforms used in mental health interventions to create a more client-centered and engaging experience.

### Social Robots and Well-Being

To address engagement and motivational difficulties with mobile mental health interventions, researchers have begun to explore the possibility of using animated characters and social robots as personalized social companions to deliver or supplement behavioral interventions [25,26]. Socially assistive robots (SARs) are robotic technology platforms with audio, visual, and movement capabilities. Their purpose is to create friendly and effective interaction with a human user with the additional aim of giving assistance to the user and achieving measurable progress in quality of life, often related to motivation, rehabilitation, or learning [27]. SARs are embodied, taking up physical space in the world and not merely existing on a screen, and can use audio and/or closed captioning to converse socially with humans, depending on their design [28]. It is important to note that SARs are both platforms for interventions and also interventions in and of themselves; they can learn and engage socially with individuals while also presenting interventions to users similar to mobile apps (eg, skills training, health tracking). They can engage users across multiple sensory options, most often including sound, sight, and touch, which can create multiple modalities for the delivery of content or interactions, depending on user preferences or personal physical abilities [29]. Given their multiple abilities, SARs may potentially integrate traditional app- and telehealth-related supports with an interactive social companion, providing a more engaging and responsive platform for users.

Although research with SARs is still in its early stages, preliminary research has reported positive participant responses to SARs assisting in physical health interventions related to increasing exercise with the elderly [30], improved cardiac rehabilitation through self-reported usefulness of SARs to assist in completion of rehabilitation tasks [31], and improved medication management through medication reminders by an SAR [32]. These specific robots were created to serve as embedded reinforcers of tasks, health behaviors, and prosocial interactions and are used across a wide range of conditions. One study found that SARs may assist with weight management, motivation, and self-monitoring strategies, with engagement sustained beyond what has been found with the same treatment delivered passively online [26].

There is limited evidence that SARs can assist with mental health and well-being interventions in pediatric populations by providing comfort or coaching [33,34]. Additionally, a 2013 meta-analysis focusing on the psychological outcomes of robot-enhanced therapies suggests that social robots could be used as a complementary tool in therapy for specific populations, particularly with children [35]. The function of social robots in adult populations is different and should be studied separately; children and adolescents often respond differently to robots, and the focus should be on developmental and skill-based learning (eg, to support children’s play [36] or assist adolescents with autism [37]). However, these platforms and their socially responsive capabilities can be modified to assist adults with their mental and behavioral health goals, given the dual nature of social robots to create personalized, affective relationships with users and assist in setting, tracking, and supporting users in meeting specific goals [38]. Unfortunately, little is known about the nature of social robots and their potential use in assisting in the psychological well-being of adult populations.

Social robots have been used in research with children, usually involving children with developmental disabilities [39]. However, the function of social robots is different in pediatric versus adult populations. Research with children focuses on robots as models of appropriate behavior or physical helpers with manual tasks [40]. There is more variability in the functionality of social robots in research with adults, thus we restricted our search to adult populations.

There is room for expansion in the use of SARs in mental health research with adults and how the research into robots and mental health has developed in recent years given the fast-paced nature of robotics. Reviewing the existing empirical literature on the use of social robots for mental health interventions is essential to determine the current state of the research and make suggestions regarding future areas for investigation. This systematic review of the literature attempts to synthesize studies using social robots to affect mental health and psychological well-being outcomes, identify the current strengths and weaknesses in the research, and suggest opportunities for growth and exploration.

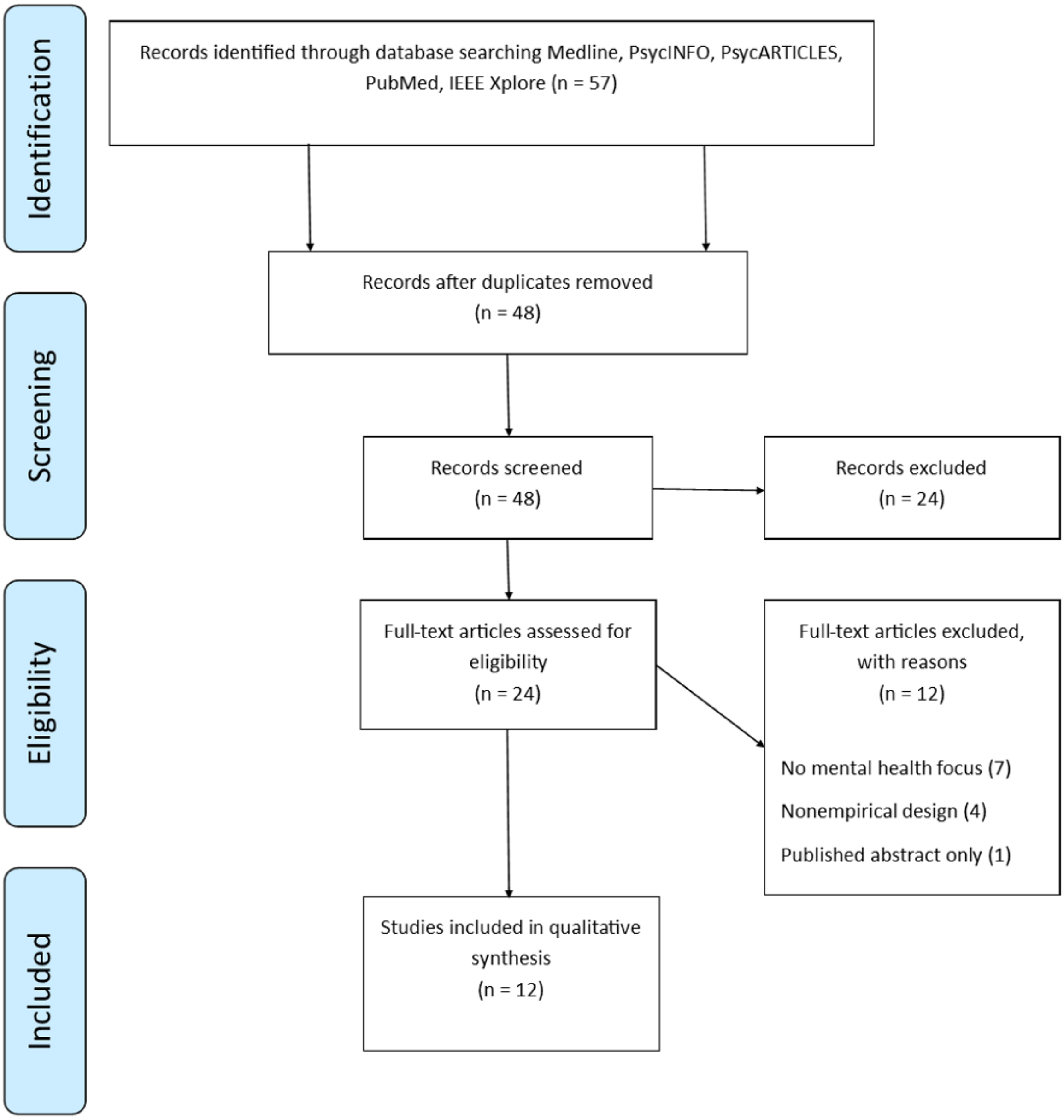
## Methods

This study is a systematic synthesis of the literature from the past 10 years that examines the use of social robots in mental health and psychological well-being. The fields of robotics and artificial intelligence are rapidly changing and this review is meant to reflect the current research in this area. We sought to answer the question: How have social robots been used to enhance mental health services for adults? A search was conducted on June 20, 2018, in the databases Medline, PsycINFO, PsycARTICLES, PubMed, and IEEE Xplore. We used several search term combinations to search titles, abstracts, keywords, and text in articles: social robot* + mental health, social robot* + counseling, social robot* + therapy, social robot* + psychotherapy, socially assistive robot* + mental health, socially assistive robot* + counseling, socially assistive robot* + therapy, socially assistive robot* + psychotherapy. Search terms were developed in consultation with coauthors and a research librarian. In our search, we excluded studies published in languages other than English, studies published prior to 2008, and studies that focused on pediatric populations. This search yielded 48 articles in total for abstract review after removing duplicates (n=9). Exact search strings for each database are included in Multimedia Appendix 1.

The authors used Covidence.org to organize the review and conduct blinded abstract and full-text reviews. This review is not formally registered. Two authors independently reviewed all abstracts and met to come to consensus on the inclusion or exclusion of articles in conflict. The inclusion criteria were that the article was an empirical study involving data collected on the direct interactions between a human participant (aged 18 years or older) and a social robot (an embodied robotic platform meant to form an assistive or affective connection with users) and that the authors explicitly stated their study focused on a mental health treatment population (eg, psychiatric patients) or used a mental health–focused intervention (eg, motivational interviewing). Included studies had to report on one or more mental health or psychological well-being outcomes with data collected to measure the robot's possible relationship with the mood, psychological welfare, or comfort of users. Although our search process did not include the term psychological well-being, it became clear in our review process that articles that explicitly aimed to examine mental health actually measured aspects of well-being rather than specific mental health constructs. Thus, we expanded our inclusion criteria to include measurement of a psychological well-being outcome. Twenty-four articles were excluded in the abstract review process, detailed in Figure 1. There was a discrepancy on rating abstracts with 2 articles that was resolved through consensus agreement, indicating good interrater reliability. This left 24 articles for full-text review. The same two authors independently reviewed all 24 full-text articles and met to resolve any conflicts by consensus agreement. Twelve articles were excluded in the full-text review phase because they did not use an empirical design (eg, it was a theoretical or commentary paper) or the focus of the study was not on mental health or well-being (eg, mental health was not a primary outcome). Twelve studies remained for inclusion in our final review.

Data were extracted from the 12 articles included in the final review. Specifically, detailed information about the sample size and characteristics of the population, study design, mental health or well-being outcome and measurement, robot used, intervention implemented, study findings, and possible biases were all recorded. Details about definitions of mental health or well-being outcome and purpose of the intervention were also identified. Studies were assessed for methodological quality, but quality assessments were not used to exclude any studies. Instead, quality assessments served to identify consistent weaknesses across studies. Contact with individual study authors was ultimately not deemed necessary to extract needed information from the included studies.

**Figure 1.** Preferred Reporting of Systematic Reviews Search and Review diagram.




## Results

### Summary

The 12 studies included in our review used 5 different social robots (Paro, NAO, CRECA, Betty, and Haptic Creature) constituting 3 major areas of robotic applications for mental health developed from our review: comfort/companionship, stress reduction, and motivation. The studies were published between 2010-2018 in a variety of peer-reviewed journals; 4 studies were published conference papers. Sample sizes for the studies ranged from 2 to 248; 7 of the 12 studies were conducted in elderly populations in nursing home settings, 2 were conducted with college students, 2 with hospital staff, and 1 with women aged 19 to 45 years recruited from the community. A minority of studies (n=3) focused on participants who were aged 45 years or less. Three studies were not conducted with a clinical mental health sample but from participants recruited from the community or from local colleges. These studies were included in the review because they measured mental health or well-being outcomes such as self-reported anxiety or stress reduction. Table 1 describes the studies, samples, interventions, and main findings with respect to the effects of social robots on psychological outcomes collected in our review.

**Table 1.** Main evidence from systematic review.

| First author, year | Study design | Sample size and characteristics | Robot | Mental health or well-being outcome | Intervention | Main findings |
|---|---|---|---|---|---|---|
| Bemelmans, 2015 [41] | Pre-post (single session) | 71 nursing home residents with dementia (age range not reported) | Paro | IPPA[a] score and mood via COOP/WONCA[b] chart recorded by care provider | Quasi-experimental time series study: 15-minute interaction with Paro while experiencing unrest or negative mood | Significant positive effect on mood and IPPA score |
| Galvão Gomes da Silva, 2018 [42] | Pre-post (single session) | 20 psychology students (aged 18 to 62 years, majority under 25 years) | NAO | 24 open-ended self-reported items in questionnaire assessing motivation for exercise (author created) | Two lab sessions of motivational interviewing for exercise with NAO (1-week interval between sessions) | • Positive appraisals of robot as nonjudgmental<br>• Increased “change talk” in participants<br>• Increased motivation to exercise |
| Kurashige, 2017 [43] | Pre-post (single session) | 12 male students aged 21 to 23 years (mean age not reported) | CRECA[c] | Author-created self-report items (15) on conversational flow, perceived trust/reliability in CRECA, and stress reduction | Motivational interviewing session with nodding or not nodding CRECA around stress management | • Positive appraisal of dialogue with nodding CRECA<br>• Self-reported reduction in anxiety |
| Lane, 2016 [44] | Pre-post | 106 VA[d] community living center elderly patients (aged 58 to 97 years, mean age 80 years) | Paro | Care staff observed behaviors and mood on researcher-created tracking sheet across 3 time periods (baseline, Paro treatment, posttreatment) | Veteran was actively presented with or observed to be actively using Paro for a minimum of 5 minutes | • Observed that Paro reduced negative behavior and mood states<br>• Observed that Paro induced increases in indicators of positive mood states |
| Loi, 2017 [45] | Pre-post | 45-bed unit for younger adults with neuropsychiatric conditions, (residents < 65 years, mean age not reported) | Betty | Staff completed a pre- and post-SARs[e] questionnaire regarding patient well-being, enjoyment, and quality of life (items based on technology acceptance model) | Betty was present at the facility for 12 weeks; engaged with residents via conversations, music, relaxation exercises, and games | Staff reported that Betty was helpful to patients by being comforting, relaxing, and improving the well-being of residents |
| Moyle, 2018 [46] | Cluster randomized RCT[f] | Residents with dementia in a long-term care facility (mean age not reported) | Paro | Staff reported patient comfort and well-being (qualitative interview) | Three 15 minute interactions were observed between Paro and elderly residents within 3 treatment groups: Paro, plush toy, or usual care | Staff indicated there were benefits to using Paro as a companion to elderly patients, although Paro did not comfort all residents |
| Šabanović, 2013 [47] | Pre-post | 10 nursing home residents with dementia (ages not reported) | Paro | Researcher videotaped and coded interactions based on positive engagement with others | Residents interacted with Paro over 7 weekly sessions | Observed an increase in prosocial interaction between residents |
| Sefidgar, 2016 [48] | Pre-post (single session, within-subject design) | 38 women aged 19 to 45 years, mean age 23.8 years | Haptic Creature | Self-reports on the SAM[g]; STAI-6[h] | Interaction with Haptic Creature on lap, compared with nonmoving stuffed animal replica | • Biometric indicators of relaxation related to heart and respiration rates significantly decreased relative to stroking a nonbreathing replica<br>• Participants reported feeling calmer and happier |

| First author, year | Study design | Sample size and characteristics | Robot | Mental health or well-being outcome | Intervention | Main findings |
|---|---|---|---|---|---|---|
| Valentí Soler, 2015 [49] | Pre-post | 211 nursing home patients with dementia, 37 at day care facility (total n=248; age range 58 to 100 years, mean age 84.7 years | Paro, NAO | Staff reported on the Apathy Inventory and QUALID[i] scale | Comparing interactions with Paro, NAO, and live dog over 3 months | • Apathy and irritability improved for NAO and Paro groups<br>• Quality of life improved for Paro group |
| Wada, 2010 [50] | Pre-post | 2 elderly individuals and 1 caregiver, age not reported | Paro | Researcher observed emotional responses and behaviors (ie, smiling) | Caregivers engaged in a manual-assisted 30-minute interaction between residents and Paro (4 sessions) | Observed-recorded increase in positive behaviors in 1 participant (smiling, laughing), no significant change in other participant |
| Wada, 2012 [51] | Pre-post | 12 elderly participants (mean age 86.8 years) and 9 caregivers (mean age 28.1 years) | Paro | Observation sheet recording participant behaviors and emotional reactions (researcher-recorded) | Manual-assisted interaction with Paro; observed before caregiver used manual and after caregiver used manual | The manualized Paro interaction increased contentment and positive social interactions |
| Wada, 2014 [52] | Pre-post | 64 elderly individuals in 7 elder-care facilities (mean age 86.5 years) | Paro | Observation sheet recording perceived participant behaviors and mood (anxiety, depression, aggression) | Manual-assisted interaction with Paro over 5 months | Following Paro interactions, caregivers observed decreases in perceived anxiety, depression, or aggression in 25 residents (39%) |

[a]IPPA: Individually Prioritized Problems Assessment.

[b]COOP/WONCA: Primary Care Cooperative Information Project/World Organization of Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians.

[c]CRECA: Contextual Respectful Counseling Agent.

[d]VA: US Department of Veterans Affairs.

[e]SAR: socially assistive robot.

[f]RCT: randomized controlled trial.

[g]SAM: Self-Assessment Manikin.

[h]STAI-6: State-Trait Anxiety Inventory.

[i]QUALID: Quality of Life in Late Stage Dementia.

## Robotic Devices

A majority of studies (n=8) used Paro, a robot that resembles a baby harp seal, with 2 studies using NAO, and 1 study using CRECA (Contextual Respectful Counseling Agent), Haptic Creature, and Betty. In Figure 2 and Table 2, pictorial and text descriptions of the 5 social robots are provided. Overall, the appearance of the robots used matched their purpose, with Paro and the Haptic Creature resembling animals, as the researchers aimed to use audio, visual, and tactile sensors to mimic animal-assisted therapy. The humanoid SARs (CRECA, Betty, and NAO) had audio, visual, and tactile sensors as well but also used additional sensors to communicate verbally with users to provide interactions related to relaxation or mental health treatment (ie, counseling, motivational interviewing). Although all the robots were capable of limited movement, only NAO was able to walk and assume a standing position if needed. The weight of the robots varied greatly, with the heaviest weighing 14 lbs (Betty) and the lightest weighing 6 lbs (Paro; information on height/weight of CRECA not available).

**Figure 2.** Social robots used in reviewed articles: (a) Paro, AIST [41]; (b) Haptic Creature [48], photo by Martin Dee; (c) NAO, Aldebaran Robotics; (d) Betty [45]; and (e) CRECA [43].

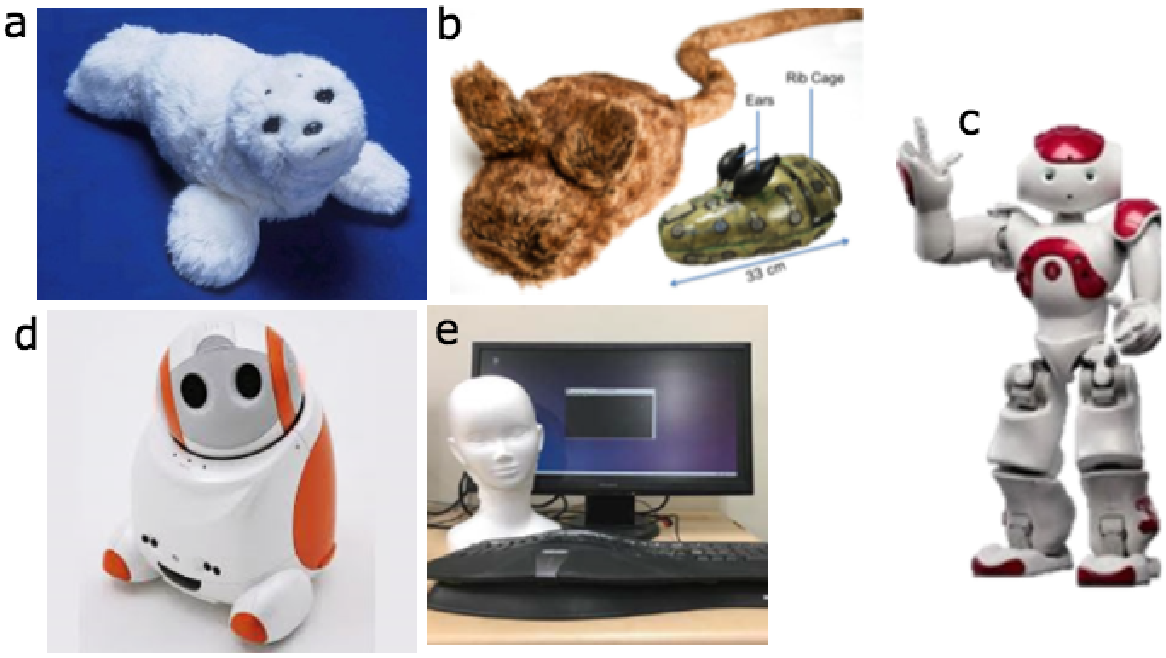


**Table 2.** Description of social robots used in reviewed articles.

| Robot | Physical appearance and specs | Sensors | User interactivity | Purpose |
|---|---|---|---|---|
| Paro | Paro is a robotic harp seal, weighing 6 lbs and 22.4 inches long. Paro can be recharged through its “pacifier” battery charger. Its fur is removable, washable, fluffy, and antibacterial. The US Food and Drug Administration has classified Paro as a “biofeedback medical device,” and the platform is not programmable by external users. | Has five kinds of sensors: tactile, light, audition, temperature, and posture sensors, with which it can perceive people and its environment. | He can sense when being touched by its tactile sensor, or when being held by a posture sensor. Can also recognize the direction of voice and words such as its name, greetings, and praise with its audio sensor. His voice imitates a harp seal. | Paro is meant to provide similar comfort as animal therapy for patients in facilities where live animals may present treatment or logistical difficulties. Paro may be used for comfort, companionship, or stress reduction. |
| Betty | Betty is an individualized, socially-assistive robot, with the technological name Partner Personal Robot PaPeRo. Betty is 15.35 inches tall and weighs about 14 lbs. Betty is programmable by users external to the company. | Betty has audio, touch, movement, and visual sensors; specifically, it is equipped with a camera, microphone, a touchpad, and speakers. | Voice recognition is the primary modality for interacting with Betty. The robot can also make human-like gestures, has voice recognition capabilities, is mobile, and can be programmed with a person’s preferences (eg, books, games, or music). | Betty may be used for motivation, entertainment, or companionship. The robot is meant to provide human-like interactions and reciprocal engagement, while also providing a calming effect for users. |
| NAO | NAO is brightly colored with large eyes and humanoid appearance. NAO stands at 22.8 inches tall and weighs 12.1 lbs. Its default walking speed is 0.2 mph. The robot is fully programmable. | It has multiple sensors for touch, sound, speech, and visual recognition. NAO is also capable of movement, with both fall and fall recovery capabilities. | NAO interacts with users via an audio system, often with accompanying movements and lights. It has speech recognition and dialogue with NAO is available in 20 languages. | It has been used in research with children who have developmental disorders or disabilities. NAO is also used for motivation or companionship |
| Haptic Creature | The Haptic Creature is a comfort robot and was designed based on human-animal interaction models. It is characterized as an expressive animatronic lap-pet (size of a large cat). It is 12.9 inches long and weighs 5.5 lbs. The platform utilizes custom programming that may be available to external programmers upon request. | Includes a 30-item touch dictionary developed from social psychology and human-animal interaction literature. It perceives movement and touch, responding with ear stiffness, modulated breathing, and vibrotactile purring. | Users interact with the Haptic Creature solely through touch, with the robot responding with movement and visual cues to mimic relaxed breathing. | Through touch, it promotes emotional interaction with the user with the aims to reduce anxiety similar to animal assisted therapy. It can also be used for comfort or stress reduction. |
| CRECA | CRECA stands for “Context Respectful Counseling Agent” and works in conjunction with an on-screen counseling agent avatar. The platform utilizes custom programming that may be available to external programmers upon request | This robot is connected to a computer and microphone to perform speech functions using natural language processing. It can also perform nodding movements. | It can converse with the users, respond to client verbalizations with prompts for continued discussion, and nod its head to validate the user’s responses. | CRECA primarily serves as an educational or motivational robot that can mimic the verbal and non-verbal interactions between counselors and clients |

## Study Interventions

The reviewed studies had 2 types of interventions and varied in robotic-interaction frequency. Four studies conducted a single-time lab research session to assess their SARs [41-43,48], while the majority (n=8) embedded their SARS within a facility over a specified period of time to evaluate the impact of their SARs on residents. None of the studies employed a robot within the personal home of a participant. Although inclusion criteria were meant to capture SAR research with a mental health focus and on persons with identified mental health issues, no studies were conducted specifically with persons reporting psychological diagnoses.

In the reviewed studies, the SAR was used as the intervention, and participants were assessed to see if interacting with the robot resulted in changes in well-being or mental health. In addition, although all of the studies used at least one social robot as the main platform for a well-being or mental health intervention, one [49] used two robots in order to compare either a live animal, Paro, or NAO as the mode of delivering comfort to nursing home residents with dementia. In addition, 3 studies [43,46,48] used less responsive comparisons as controls for their SARS, including a plush animal comparison (n=2) or a nonnodding CRECA (n=1). A minority of studies (n=2) used an SAR as a platform to conduct motivational interviewing sessions with social robots, one focused on motivation to exercise using NAO and the other focused on stress reduction using CRECA [43,48]. The remaining 10 research studies focused on the social robot as a means of comforting, increasing positive emotions, or providing companionship to participants, with 9 of these using the robots within the context of elderly or long-term care settings.

### Study Design and Measurement

Our review revealed that 6 of the studies used clinical care staff to observe and measure outcomes related to mood and mood-related behavioral changes, 3 used self-report data [42,43,48], and 3 had researchers use tracking sheets [44,46,50] to report on mood and mental health outcomes. These methods may be an artifact of the populations studied, due to the majority of studies involving persons with cognitive issues who would make self-report challenging. In addition, only 4 studies [41,46,48,49] used nonresearch-created questionnaires, using previously validated measurements such as the Apathy Inventory and Quality of Life in Late Stage Dementia scale. Although one study [46] used an RCT (randomized controlled trial) study design, most of the studies (n=11) used a pre-post design, measuring changes in psychological well-being or intervention impact before and after interacting with a SAR.

### Mental Health Outcomes

Overall, results regarding the impact of social robot–delivered mental health interventions and interactions ranged from generally positive to mixed. The majority of studies focused on symptom reduction related to mood and positive quality of life changes after robot interactions. The majority (n=11) reported positive increases in mood, comfort, or stress reduction following the social robot interventions, although 2 [46,50] showed mixed results on whether Paro comforted elder-care residents. For the 7 studies that focused on elderly and dementia populations, outcomes included observed aggression, contentment, anxiety, and depression [41,44,45,49-52]. However, nearly all studies had a main goal of assessing feasibility or usability of the social robot in a given population. Since mental health outcomes were secondary and these studies were pilot studies, psychometrically validated measures were seldom used and instead measures were frequently created for the particular study.

Two studies reported on functional outcomes related to quality of life. These studies focused specifically on physical health and well-being, with one reporting a generally positive impact on exercise [42] and one reporting a reduction in physical indicators of stress and an increase in self-reported mood after interaction with a social robot [41]. Finally, 2 studies assessed the ways in which social robots facilitate positive social interactions, with one reporting increased social interaction among participants in a nursing home and their caregivers following an interaction with Paro [51] and the other noting that social interactions among residents in a nursing home increased after Paro was integrated into the facility [47].

## Discussion

### Principal Findings

Overall, our review revealed the nascent nature of mental health research with social robots. Although there is a rising interest in using social robots in psychological interventions, there is still a very modest research base examining this application. Our 12 reviewed studies included the use of 5 distinct social robots to influence various mental health or well-being outcomes. The majority focused on providing comfort and companionship to study participants (Paro, Haptic Creature). A minority of studies used SARs to implement a specific intervention (eg, motivational interviewing with NAO and nodding CRECA). The impact of social robot–delivered mental health interventions and interactions ranged from generally positive to mixed, with some studies finding positive changes in mood and quality of life after robot interactions.

This review suggests that existing studies of the potential impact of SARs on psychological well-being are limited in generalizability, scope, and measurement. Specifically, nearly all of the studies conducted in this area have occurred in elderly care facilities or laboratory settings, with a bimodal distribution of participants ranging from quite young (under 24) to elderly populations within nursing home facilities (often over 65 years). Our findings are consistent with a previous review of SAR use in care of the elderly [35]), which found studies in this area to be methodologically limited, so much so that even optimistic findings required additional replication prior to making clear conclusions about SAR effectiveness. This also relates to our finding that the majority of reviewed studies used clinical care or researcher-reported outcomes. Specifically, since multiple studies were conducted with elderly samples suffering from dementia or neuropsychiatric conditions, observational data may have been considered most feasible and more valid than self-report data. Similarly, surveying caretakers of impaired elderly persons may have been more feasible than surveying the patients themselves. The lack of validated scales used in the studies underscores the nascent state of this field and the need for further and more structured research.

Social robot–delivered interventions may constitute a promising treatment for chronic conditions and health management needs in elderly populations based on the findings of studies included in this review. In particular, the available data indicate that Paro may be a useful tool for increasing socialization and decreasing aggression in dementia populations. However, there is insufficient evidence that this finding can be generalized to other populations or even to nursing home residents without dementia. In addition, some studies identified were conference paper proceedings, which highlights a limitation of the research. Conference papers are not necessarily peer reviewed and may indicate a higher risk of bias.

### Limitations

Generalizability of these findings is also limited by the study characteristics. Many of the studies had very small sample sizes, which also limits the generalizability of findings. Moreover, included studies frequently had very brief interventions with simple pre-post study designs, which might make it difficult to assess differences in pre and post data and preclude conclusions about the efficacy of the interventions. We did not include specific mental health diagnoses in our searches, and therefore some studies using SARs to target very specific diagnoses, rather than mental health more generally, may have been excluded. In addition, we did not include papers that focused only on the development of a particular SAR because we only included studies that had participants. There is also a lack of definitional clarity in the concept of social robots in general, and this may lead to different conceptualizations of SARs in the literature.

There is a clear need for greater testing and programming of robots to assist with patient care within the home. Although many of the research studies reported a future aim of using social robots within the home for mental health support and interventions, no study had currently begun this level of at-home technology testing. Without such testing, it is unclear whether an at-home mental health companion, coach, or motivator would have a strong positive effect on the quality of life of patients managing their depression, anxiety, or other psychiatric conditions where health happens most—in the home. Choosing the appropriate SAR to use for future research should entail a review of not only this research but also the robot characteristics necessary for the user. For instance, many of the robots reviewed weighed more than 10 lbs, which can be contraindicated with certain populations should they need to lift the SAR. Such considerations are imperative in order to align an intervention's purpose, user, and chosen robotic platform.

Methodologically, there is room for improving and extending the current research base of social robots in mental health interventions. The results of this review show that nearly all studies in this area are preliminary or pilot studies, and few include validated measures of mental health outcomes. Often, the primary aim of such research is usability, feasibility, and acceptability of the social robots with mental health measures secondary to this main goal. Mental health outcomes were often vaguely defined as the observed reduction of anxiety or negative mood symptoms. Expanded use of self-report questionnaires or clinician-administered measures with psychometric validation is indicated. In addition, outcomes should be aligned with what patients might care about in treatment outcomes, which may include an increase in functional abilities, social interactions, or self-reported quality of life. Nearly all of the studies reviewed used single-session robot interactions that may make tracking changes in mental health difficult. As previously stated, many studies included participants with cognitive difficulties, which might make self-report assessments difficult or impossible, hence their use of observed mental health outcomes. However, if the purpose of social robot–delivered treatments is to capitalize on the functionality of such robots—neutral, asynchronous, always available, and capable of personalized tracking and feedback—such testing and adequate measurement is essential to the creation of patient-centered social robots. Because of the heterogeneity in how outcomes were reported, we could not perform a meta-analysis or draw conclusions about possible biases at work across studies.

## Practical Implications and Future Research

After reviewing the existing research on social robots and mental health, it is clear that there are ample opportunities to test and measure the ways social robots could be useful adjuncts to mental health treatments for various adult populations. Current research has used mental health outcomes as a secondary focus, and future research should explore the potential benefits of SARs in specific clinical populations with difficulty accessing care. Examples of such populations might include veterans living with chronic pain in rural areas, individuals with mental health needs who cannot make appointments during regular business hours, individuals with transportation issues, or individuals who feel stigmatized in traditional mental health care settings. Exploration of motivational, companionship, and social facilitation functions of SARs were assessed in a minority of studies. In Table 3, we highlight further recommendations and areas of consideration for future research into the mental health and clinical applications of social robots.

**Table 3.** Considerations and recommendations for future research.

| Research considerations | Recommendations |
|---|---|
| Internal validity | • Improve upon and expand methods beyond pilot studies<br>• Use validated mental health outcome measures when advancing beyond pilot feasibility studies<br>• Account for potential mediators between socially assistive robot interactions and mental health outcomes, such as usability or technology issues |
| External validity/generalizability | • Expand beyond dementia and developmentally disordered populations to include a range of ages and diagnoses (with special attention to those who may not currently have access to needed care)<br>• Explore use of socially assistive robots across different settings, from medical facilities to at-home robots |
| Inclusion of theory | • Use existing literature on human-robot interactions to account for study aims and interventions design decisions<br>• Embed psychological theory into future research—such as object relations—to explore individual mental health outcomes and reactions and perceived efficacy of socially assistive robots |
| Dissemination and translation | • Expand future research to robots that can engage in more human-like social interaction<br>• Consider close, multidisciplinary collaborations (eg, between clinicians, researchers, and robotics programmers) to allow for iterative and responsive intervention development |
| Cultural concerns | • Investigate the impact of sociocultural beliefs and differences related to technology comfort, linguistic challenges, and interest in socially assistive robots for mental health<br>• Focus on specific mental health populations that might be uniquely suited to benefit from socially assistive robots |

## Conclusions

Our review sought to examine how social robots have been used to influence mental health in general, and a possible limitation of our review is that we chose not to include specific mental health terms (eg, depression) in our search process. In addition, since our search, which focused on mental health, yielded studies that mostly assessed aspects of well-being, it is possible that including multiple more specific mental health terms might have yielded different results.

Overall, better integrating and expanding on the mental health implications of social robots will clearly complement the ongoing drive in the field of psychology to better assist clients at home with supportive exercises, goal tracking, and an asynchronous care option. Although our review revealed that the use of SARs in mental health research is not yet widespread, new robots and programming are constantly changing, adapting, and expanding. The use of SARs in mental health research and mental health interventions is nascent and has thus far been restricted to specific populations with limited measurement and scope. There is an abundance of opportunity in this area for growth, expansion, and exploration to triangulate SARs usability and efficacy data as the next step in advancing this field.


## Acknowledgments

This material is the result of work supported with resources and the use of facilities at the Edith Nourse Rogers Memorial Veterans Hospital. The findings and interpretations of the data expressed in the article are the sole responsibilities of the authors and do not necessarily represent the views of the Department of Veterans Affairs.


## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Search strings.

[DOCX File, 14KB - jmir_v21i7e13322_app1.docx ]

## References


1. Lawrence D, Kisely S. Inequalities in healthcare provision for people with severe mental illness. J Psychopharmacol 2010 Nov;24(4 Suppl):61-68 [FREE Full text] [doi: 10.1177/1359786810382058] [Medline: 20923921]
2. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. Lancet 2007 Sep;370(9590):878-889. [doi: 10.1016/s0140-6736(07)61239-2]
3. Goddard M, Smith P. Equity of access to health care services: theory and evidence from the UK. Soc Sci Med 2001 Nov;53(9):1149-1162. [doi: 10.1016/s0277-9536(00)00415-9]
4. West A, Weeks WB. Physical and mental health and access to care among nonmetropolitan veterans health administration patients younger than 65 years. J Rural Health 2006 Jan;22(1):9-16. [doi: 10.1111/j.1748-0361.2006.00014.x]
5. Dahl T, Boulos M. Robots in health and social care: a complementary technology to home care and telehealthcare? Robotics 2013 Dec 30;3(1):1-21. [doi: 10.3390/robotics3010001]
6. Peeters JM, Wiegers TA, Friele RD. How technology in care at home affects patient self-care and self-management: a scoping review. Int J Environ Res Public Health 2013 Oct 29;10(11):5541-5564 [FREE Full text] [doi: 10.3390/ijerph10115541] [Medline: 24173139]
7. Heapy AA, Higgins DM, Goulet JL, LaChappelle KM, Driscoll MA, Czlapinski RA, et al. Interactive voice response-based self-management for chronic back pain: the COPES noninferiority randomized trial. JAMA Intern Med 2017 Dec 01;177(6):765-773 [FREE Full text] [doi: 10.1001/jamainternmed.2017.0223] [Medline: 28384682]
8. Comer JS. Introduction to the special series: applying new technologies to extend the scope and accessibility of mental health care. Cogn Behav Pract 2015 Aug;22(3):253-257. [doi: 10.1016/j.cbpra.2015.04.002]
9. Wang J. Perceived barriers to mental health service use among individuals with mental disorders in the Canadian general population. Med Care 2006 Feb;44(2):192-195. [Medline: 16434920]
10. Levin M, Stocke K, Pierce B, Levin C. Do college students use online self-help? A survey of intentions and use of mental health resources. J Coll Stu Psychother 2017 Oct 09;32(3):181-198. [doi: 10.1080/87568225.2017.1366283]
11. Andrade L, Alonso J, Mneimneh Z, Wells J, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychological medicine 2014;44(6):1303-1317. [Medline: 23931656]
12. Kuhn E, Weiss B, Taylor K, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I Coach: a description and clinician perceptions of a mobile app for cognitive behavioral therapy for insomnia. J Clin Sleep Med 2016 Apr 15;12(4):597-606 [FREE Full text] [doi: 10.5664/jcsm.5700] [Medline: 26888586]
13. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]

14. Ryan C, Lewis J. US Department of Commerce, Economics and Statistics Administration, US Census Bureau. 2015. Computer and internet use in the United States URL: https://www.census.gov/content/dam/Census/library/publications/2017/acs/acs-37.pdf [accessed 2019-06-24]
15. Cahill S, Macijauskiene J, Nygård A, Faulkner J, Hagen I. Technology in dementia care. Technol Disabil 2007;19(2):55-60 [FREE Full text]
16. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017 Aug;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
17. Torous J, Nicholas J, Larsen M, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
18. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: a novel mhealth system for supporting self-care in management of complex and chronic conditions. JMIR Mhealth Uhealth 2013 Jul 11;1(2):e10 [FREE Full text] [doi: 10.2196/mhealth.2391] [Medline: 25100682]
19. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and Web-based text messaging in mental health. J Med Internet Res 2016 Dec 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
20. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018 May;9:797. [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
21. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Dec;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
22. Short C, Rebar A, Plotnikoff R, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. Health Psychol Rev 2015;17(1):32-33 [FREE Full text]
23. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. mHealth 2018 Mar;4:6. [doi: 10.21037/mhealth.2018.03.02]
24. 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]
25. 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]
26. Kidd C, Breazeal C. Robots at home: understanding long-term human-robot interaction. 2008 Presented at: IEEE/RSJ International Conference on Intelligent Robots and Systems; 2008; Nice. [doi: 10.1109/iros.2008.4651113]
27. Feil-Seifer D, Mataric M. Defining socially assistive robotics. 2005 Presented at: 9th International Conference on Rehabilitation Robotics; 2005; Chicago p. 465-468.
28. Mollahosseini A, Abdollahi H, Sweeny TD, Cole R, Mahoor MH. Role of embodiment and presence in human perception of robots' facial cues. Int J Hum Comput Stud 2018 Aug;116(116):25-39 [FREE Full text] [doi: 10.1016/j.ijhcs.2018.04.005]
29. Causo A, Vo G, Chen I, Yeo S. Design of robots used as education companion and tutor. Robotics Mechatronics 2015:75-84. [doi: 10.1007/978-3-319-22368-1_8]
30. Fasola J, Mataric M. A socially assistive robot exercise coach for the elderly. J Hum Robot Interact 2013;2(2):1. [doi: 10.5898/jhri.2.2.fasola]
31. Kyong IK, Freedman S, Mataric M, Cunningham M, Lopez B. A hands-off physical therapy assistance robot for cardiac patients. 2005 Presented at: 9th International Conference on Rehabilitation Robotics; 2005; Chicago. [doi: 10.1109/icorr.2005.1501114]
32. Smarr C, Prakash A, Beer JM, Mitzner TL, Kemp CC, Rogers WA. Older adults' preferences for and acceptance of robot assistance for everyday living tasks. Proc Hum Factors Ergon Soc 2012 Sep;56(1):153-157 [FREE Full text] [doi: 10.1177/1071181312561009] [Medline: 25284971]
33. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Effectiveness of robot paro in intramural psychogeriatric care: a multicenter quasi-experimental study. J Am Med Dir Assoc 2015 Nov 01;16(11):946-950. [doi: 10.1016/j.jamda.2015.05.007] [Medline: 26115817]
34. Blanson Henkemans OA, Bierman B, Janssen J, Neerincx M, Looije R, van der Bosch H, et al. Using a robot to personalise health education for children with diabetes type 1: a pilot study. Patient Educ Couns 2013 Aug;92(2):174-181. [doi: 10.1016/j.pec.2013.04.012] [Medline: 23684366]
35. Costescu CA, Vanderborght B, David DO. The effects of robot-enhanced psychotherapy: a meta-analysis. Rev Gen Psychol 2014;18(2):127-136. [doi: 10.1037/gpr0000007]
36. Interactive robotic social mediators as companions. URL: http://www.iromec.org [accessed 2019-06-24]
37. Vanderborght B, Simut R, Saldien J, Pop C, Rusu AS, Pintea S, et al. Using the social robot probo as a social story telling agent for children with ASD. Interactive Stud 2012;13(3):348-372. [doi: 10.1075/is.13.3.02van]

38. Rabbitt S, Kazdin AE, Scassellati B. Integrating socially assistive robotics into mental healthcare interventions: applications and recommendations for expanded use. Clin Psychol Rev 2015 Feb;35:35-46. [doi: 10.1016/j.cpr.2014.07.001] [Medline: 25462112]
39. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. Annu Rev Biomed Eng 2012 Aug;14:275-294. [doi: 10.1146/annurev-bioeng-071811-150036] [Medline: 22577778]
40. Kim ES, Berkovits LD, Bernier EP, Leyzberg D, Shic F, Paul R, et al. Social robots as embedded reinforcers of social behavior in children with autism. J Autism Dev Disord 2012 Oct 31;43(5):1038-1049. [doi: 10.1007/s10803-012-1645-2]
41. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Effectiveness of robot Paro in intramural psychogeriatric care: a multicenter quasi-experimental study. J Am Med Dir Assoc 2015 Nov 01;16(11):946-950. [doi: 10.1016/j.jamda.2015.05.007] [Medline: 26115817]
42. Galvão Gomes da Silva J, Kavanagh DJ, Belpaeme T, Taylor L, Beeson K, Andrade J. Experiences of a motivational interview delivered by a robot: qualitative study. J Med Internet Res 2018 May 03;20(5):e116. [doi: 10.2196/jmir.7737]
43. Kurashige K, Tsuruta S, Sakurai E, Sakurai Y, Knauf R, Damiani E. Design of counseling robot for production by 3D printer. 2017 Presented at: 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS); 2017; Jaipur. [doi: 10.1109/sitis.2017.20]
44. Lane GW, Noronha D, Rivera A, Craig K, Yee C, Mills B, et al. Effectiveness of a social robot, “Paro,” in a VA long-term care setting. Psychol Serv 2016 Aug;13(3):292-299. [doi: 10.1037/ser0000080]
45. Loi SM, Bennett A, Pearce M, Nguyen K, Lautenschlager NT, Khosla R, et al. A pilot study exploring staff acceptability of a socially assistive robot in a residential care facility that accommodates people under 65 years old. Int Psychogeriatr 2017 Dec 04;30(7):1075-1080. [doi: 10.1017/s1041610217002617]
46. Moyle W, Bramble M, Jones C, Murfield J. Care staff perceptions of a social robot called Paro and a look-alike Plush Toy: a descriptive qualitative approach. Aging Ment Health 2016 Dec 14;22(3):330-335. [doi: 10.1080/13607863.2016.1262820]
47. Šabanović S, Bennett C, Chang W, Huber L. PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. IEEE Int Conf Rehabil Robot 2013 Jun;2013:6650427. [doi: 10.1109/ICORR.2013.6650427] [Medline: 24187245]
48. Sefidgar YS, MacLean KE, Yohanan S, Van der Loos HM, Croft EA, Garland EJ. Design and evaluation of a touch-centered calming interaction with a social robot. IEEE Trans Affective Comput 2016 Apr 1;7(2):108-121. [doi: 10.1109/taffc.2015.2457893]
49. Valentí Soler M, Agüera-Ortiz L, Olazarán Rodríguez J, Mendoza Rebolledo C, Pérez Muñoz A, Rodríguez Pérez I, et al. Social robots in advanced dementia. Front Aging Neurosci 2015 Sep 03;7:1. [doi: 10.3389/fnagi.2015.00133]
50. Wada K, Ikeda Y, Inoue K, Uehara R. "Development and preliminary evaluation of a caregiver's manual for robot therapy using the therapeutic seal robot Paro". 1 Presented at: 19th International Symposium in Robot and Human Interactive Communication; 2010; Vaireggio p. 533-538. [doi: 10.1109/ROMAN.2010.5598615]
51. Wada K, Kouzuki Y, Inoue K. Field test of caregiver's manual for robot therapy using therapeutic seal robot. Alzheimer Dementia 2012 Jul;7(4):S636-S637. [doi: 10.1016/j.jalz.2011.05.1821]
52. Wada K, Takasawa Y, Shibata T. Robot therapy at facilities for the elderly in Kanagawa prefecture: a report on the experimental result of the first month. 2014 Presented at: The 23rd IEEE International Symposium on Robot and Human Interactive Communication; 2014; Edinburgh. [doi: 10.1109/roman.2014.6926252]

## Abbreviations

**COOP/WONCA:** Primary Care Cooperative Information Project/World Organization of Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians
**CRECA:** Contextual Respectful Counseling Agent
**IPPA:** Individually Prioritized Problems Assessment
**QUALID:** Quality of Life in Late Stage Dementia
**RCT:** randomized controlled trial
**SAM:** Self-Assessment Manikin
**SAR:** socially assistive robot
**STAI-6:** State-Trait Anxiety Inventory
**VA:** US Department of Veterans Affairs

*Edited by C Eichenberg; submitted 11.01.19; peer-reviewed by R Kok, E Hermes, C Pasarelu; comments to author 05.02.19; revised version received 01.04.19; accepted 03.06.19; published 24.07.19.*

*Please cite as:*
*Scoglio AAJ, Reilly ED, Gorman JA, Drebing CE*
*Use of Social Robots in Mental Health and Well-Being Research: Systematic Review*
*J Med Internet Res 2019;21(7):e13322*
*URL: http://www.jmir.org/2019/7/e13322/*
*doi:10.2196/13322*
*PMID:31342908*

Review

# Virtual Patient Simulations in Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration

Andrzej A Kononowicz[1], MSc, PhD; Luke A Woodham[2,3], MSc; Samuel Edelbring[3,4,5], PhD; Natalia Stathakarou[3], MSc; David Davies[6], PhD; Nakul Saxena[7], PhD; Lorainne Tudor Car[8,9], MD, MSc, PhD; Jan Carlstedt-Duke[10], MD, PhD; Josip Car[11,12], MD, PhD, FRCPE, FFPH; Nabil Zary[13,14], PhD

[1]Department of Bioinformatics and Telemedicine, Jagiellonian University Medical College, Kraków, Poland
[2]Institute of Medical and Biomedical Education, St George's, University of London, London, United Kingdom
[3]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden
[4]Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
[5]Learning and Professional Development Group, School of Health Sciences, Örebro University, Örebro, Sweden
[6]Warwick Medical School, University of Warwick, Coventry, United Kingdom
[7]Health Services and Outcomes Research, National Healthcare Group, Singapore, Singapore
[8]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[9]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom
[10]President's Office, Nanyang Technological University, Singapore, Singapore
[11]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[12]Global eHealth Unit, Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom
[13]Games for Health Innovations Centre, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[14]Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

**Corresponding Author:**
Josip Car, MD, PhD, FRCPE, FFPH
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University
11 Mandalay Road, Clinical Science Building
Singapore,
Singapore
Phone: 65 6340 2480
Email: josip.car@ntu.edu.sg

## Abstract

**Background:** Virtual patients are interactive digital simulations of clinical scenarios for the purpose of health professions education. There is no current collated evidence on the effectiveness of this form of education.

**Objective:** The goal of this study was to evaluate the effectiveness of virtual patients compared with traditional education, blended with traditional education, compared with other types of digital education, and design variants of virtual patients in health professions education. The outcomes of interest were knowledge, skills, attitudes, and satisfaction.

**Methods:** We performed a systematic review on the effectiveness of virtual patient simulations in pre- and postregistration health professions education following Cochrane methodology. We searched 7 databases from the year 1990 up to September 2018. No language restrictions were applied. We included randomized controlled trials and cluster randomized trials. We independently selected studies, extracted data, and assessed risk of bias and then compared the information in pairs. We contacted study authors for additional information if necessary. All pooled analyses were based on random-effects models.

**Results:** A total of 51 trials involving 4696 participants met our inclusion criteria. Furthermore, 25 studies compared virtual patients with traditional education, 11 studies investigated virtual patients as blended learning, 5 studies compared virtual patients with different forms of digital education, and 10 studies compared different design variants. The pooled analysis of studies comparing the effect of virtual patients to traditional education showed similar results for knowledge (standardized mean difference

[SMD]=0.11, 95% CI −0.17 to 0.39, $I^2$=74%, n=927) and favored virtual patients for skills (SMD=0.90, 95% CI 0.49 to 1.32, $I^2$=88%, n=897). Studies measuring attitudes and satisfaction predominantly used surveys with item-by-item comparison. Trials comparing virtual patients with different forms of digital education and design variants were not numerous enough to give clear recommendations. Several methodological limitations in the included studies and heterogeneity contributed to a generally low quality of evidence.

**Conclusions:** Low to modest and mixed evidence suggests that when compared with traditional education, virtual patients can more effectively improve skills, and at least as effectively improve knowledge. The skills that improved were clinical reasoning, procedural skills, and a mix of procedural and team skills. We found evidence of effectiveness in both high-income and low- and middle-income countries, demonstrating the global applicability of virtual patients. Further research should explore the utility of different design variants of virtual patients.

***(J Med Internet Res 2019;21(7):e14676)*** doi:10.2196/14676

**KEYWORDS**

computer simulation; professional education; computer-assisted instruction; systematic review; meta-analysis

## Introduction

### Background

Health care education is confronted with many global challenges. Shorter hospital stays, specialization of care, higher patient safety measures, and shortage of clinical teachers all diminish the traditional opportunities for the training of health professions through direct patient contact [1,2]. Early health professions education is often dominated by theoretical presentations with insufficient connection to clinical practice [3]. The need to increase numbers and quality of the health workforce is especially visible in low-and-middle-income countries, where the need to scale up high-quality health education and introduce educational innovations is most pressing [4]. Therefore, the global medical education community is perpetually searching for methods that can be applied to improve the relevance, increase the spread, and accelerate the educational process for health professions [5].

Digital education (often referred to as e-learning) is “the act of teaching and learning by means of digital technologies” [6]. It encompasses a multitude of educational concepts, approaches, methods, and technologies. Digital health education comprises, for example, offline learning, mobile learning, serious games, or virtual reality environments. We have conducted this systematic review as part of a review series on digital health education [6-19] and focused it on the simulation modality called *virtual patients*.

Virtual patients are defined as interactive computer simulations of real-life clinical scenarios for the purpose of health professions training, education, or assessment [20]. This broad definition encompasses a variety of systems that use different technologies and address various learning needs [21]. The learner is cast into the role of a health care provider who makes decisions about the type and order of clinical information acquired, differential diagnosis, and management and follow-up of the patient. Virtual patients are hypothesized to primarily address learning needs in clinical reasoning [22,23]. However, an influence of the use of virtual patients on other educational outcomes has been reported in previous literature [21,24].

The educational use of virtual patients may be understood through experiential learning theory [25,26]. Following this theoretical model of action and reflection, virtual patients expose learners to simulated clinical experiences, providing mechanisms for information gathering and clinical decision making in a safe environment [27]. Exposing the learner to many simulated clinical scenarios supports learning diagnostic processes [28] while acquainting learners with a standardized set of clinical conditions common in the population, but rare or nonaccessible in highly specialized teaching hospitals [29].

Some concerns have been raised about educational use of virtual patients. Virtual patients should not replace but complement contact with real patients [27]. There are concerns around the use of virtual patients potentially resulting in less empathic learners [30]. The use of unfamiliar technology as part of virtual patients’ education can represent a barrier to learning, even for younger generations [31,32]. Virtual patients may also prove ineffective when technological objectives drive teaching instead of being motivated by learning needs [33].

This virtual patient simulation review has been preceded by several narrative reviews [22,34-37] and 2 systematic reviews with meta-analyses [38,39]. Our preliminary literature analysis showed that the number of studies including the term *virtual patient* or *virtual patients* has more than doubled on the MEDLINE database in comparison with available evidence provided in previous systematic reviews (February 2009 [38] and July 2010 [39]). Thus, our review will update the evidence base with studies not included in previous analyses.

### Objectives

The objective of this review was to evaluate the effectiveness of virtual patient simulation for delivering pre- and postregistration health care professions education using the following comparisons:

1. Virtual patient versus traditional education
2. Virtual patient blended learning versus traditional education
3. Virtual patient versus other types of digital education
4. Virtual patient design comparison

By traditional education, we mean all nondigital educational methods. This includes lectures, reading exercises, group discussion in classroom, and nondigital simulation as standardized patients or mannequin-based training. Virtual patient blended learning is the addition of virtual patients as a supplement to traditional education when the control intervention uses nondigital education methods only. Other types of digital education may include interventions such as video recordings, Web-based tutorials, or virtual classrooms.

We assessed the impact of virtual patient interventions on learners' knowledge, skills, attitude, and satisfaction. Our secondary objective was to assess the cost-effectiveness, patient outcomes, and adverse effects of these interventions.

# Methods

## Protocol and Registration

While conducting the review, we adhered to the Cochrane methodology [40], followed a published protocol [41], and presented results following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [42].

## Eligibility Criteria

We included randomized controlled trials (RCTs) and cluster RCTs (cRCTs). We excluded crossover trials because of the high likelihood of carryover effect.

Participants in the included studies had to be enrolled in a pre- or postregistration health-related education or training program (see glossary in Multimedia Appendix 1). This included students from disciplines such as medicine, dentistry, nursing and midwifery, medical diagnostic and treatment technology, physiotherapy and rehabilitation, and pharmacy.

This review focused on screen-based virtual patient simulations that form a computerized, dynamically unfolding representation of patient cases. A virtual patient simulation is introduced by a case description and might contain answers given by the patient, clinical data (eg, laboratory results, medical images), and descriptions of patients' signs and symptoms. Only the representations of the patient as a whole were of interest, rather than studies that focused on single parts of the body. As a matter of a policy followed in the Digital Health Education Collaboration [6] and aiming at avoiding duplication of reviews, we deliberately excluded virtual patients in 3-dimensional (3D) virtual learning environments from this study. We judged that a higher level of immersion of learners in 3D virtual environments, connected with potential technical challenges (eg, difficulties in navigating such environments, lags because of increased computational time or limited internet bandwidth), was likely to influence the educational outcomes and therefore merited a separate analysis covered already by the virtual reality review [13] of this Digital Health Education Collaboration series. We also excluded those virtual patient interventions which require nonstandard equipment (eg, haptic devices, mannequins) or those virtual patients which are human controlled (eg, simulated email correspondence or chat room conversations). We excluded studies in which virtual patients were just a small part of the intervention and those in which the influence of virtual patients was not evaluated separately.

Furthermore, 2-arm RCTs comparing virtual patients with a control group not involved in any type of subject-related learning activity were not considered eligible as previous meta-analyses have already shown a large positive effect when virtual patients were compared with no intervention [38].

We decided to introduce to the review a comparison of virtual patients blended learning with traditional education as a consequence of the discussion in the community on the need to eliminate traditional types of learning activities to make space for virtual patients. For instance, Berman et al [29] noticed that the students' subjective learning effect perceptions and satisfaction with integration were lower at universities that increased the workload of students by adding virtual patients without releasing time resources in the curriculum. As most of health professions education is conducted on campus, an integrated effect of virtual patients is possible. Blending virtual patients with traditional education is challenging and qualitatively different than a nonintervention control group comparison.

Eligible primary outcomes were students' (1) knowledge, (2) skills, (3) attitudes, and (4) satisfaction—together representing clinical competencies measured post intervention with validated or nonvalidated instruments. Secondary outcomes were (1) economic cost and cost-effectiveness, (2) patient outcomes, and (3) observed adverse effects.

## Search Methods for Identification of Studies

We searched the following 7 databases: MEDLINE (via Ovid), EMBASE (via Elsevier), The Cochrane Library (via Wiley), PsycINFO (via Ovid), Educational Resource Information Centre (ERIC; via Ovid), Cumulative Index to Nursing and Allied Health Literature (CINAHL; via EBSCO), and Web of Science Core Collection (via Thomson Reuters). We adapted the MEDLINE strategy and keywords presented in Multimedia Appendix 2 for use with each of the databases above. We searched databases from the year 1990 to September 20, 2018 to highlight recent developments and did not apply language restrictions. For all included studies, we searched references lists and conducted author and citation searches. We searched lists of references from other identified relevant systematic reviews while running our electronic searches.

## Data Collection and Analysis

### *Data Selection, Extraction, and Management*

The search results were combined in a single EndNote library (version X7; Thomson Reuters) [43]. Overall, 2 authors independently screened titles and abstracts to identify potentially eligible studies. In the next phase, full-text versions of these papers were retrieved and 2 review authors independently assessed these papers against eligibility criteria. We piloted data extraction to maximize consistency in the information extracted. Disagreements were resolved through discussion. A third review author was consulted to arbitrate when differences in opinion arose. All relevant data were extracted using a structured form in Microsoft Excel. We contacted study authors for crucial missing information, particularly if required to judge inclusion criteria and study outcomes.

### Data Items

Information was extracted from each included study on (1) the characteristics of study participants (field of study; stage of education: pre/postregistered; year of study; and country where the study was conducted and its World Bank income category: high-income/low-and-middle-income country), (2) the type of outcome measure (type of tool used to measure outcome and information on whether the tool was validated), (3) the type of virtual patient intervention (topic and language of presented virtual patient simulations; information on whether the language of virtual patient was native to the majority of participants; source of virtual patient simulations: internal/external; was the study an individual or group assignment, and in case of group assignments, the number of students in a group; whether access to virtual patient simulation was from home or in a computer laboratory; number of virtual patient cases presented; time when the virtual patients were available; and duration of use of virtual patients), and (4) the type of virtual patient system (name of the system; navigation scheme: linear, branched, and free access; control mechanism: menu-based, keyboard, or speech recognition; feedback delivery and timing; and whether video clips where included in virtual patient cases). A glossary of the terms in use in the review may be found in Multimedia Appendix 1.

### Measures of Treatment Effect

We reported the treatment effects for continuous outcomes as mean values and SDs post intervention in each intervention group, along with the number of participants and *P* values. As the studies presented data using different tools, the mean differences were recalculated into standardized mean difference (SMD). We interpreted the effect size as small (SMD=0.2), moderate (SMD=0.5), and large (SMD=0.8) effect sizes [40]. If studies had multiple arms and no clear main comparison, we compared the virtual patient intervention arm with the most common control arm, excluding the nonintervention and mixed-intervention controls. If that was impossible to decide, we selected the least active control arm. If multiple outcomes in the same category (knowledge, skills, attitudes, and outcomes) were reported, we selected the primary measure, and if that was impossible, we calculated the mean value of all measures. For papers that reported median and range for the outcomes, we converted these to mean and SD using methods described by Wan [44]. If a study did not report SD but provided CIs, we estimated SD from those using a method described in previous literature [40].

### Data Synthesis and Analysis

Owing to the significant differences between studies, we employed a random-effects model in the meta-analysis using Review Manager (version 5.3; The Nordic Cochrane Centre) [45]. We displayed the results of the meta-analysis in forest plots and evaluated heterogeneity numerically using $I^2$ statistics. For comparisons with more than 10 outcomes in the meta-analysis, we attempted a subgroup analysis. As the planned 15 subgroup analyses in the protocol [41] did not explain the heterogeneity, we visualized the outcomes using albatross plots [46]. These plots were implemented using a script created for the purpose of the study by one of the review authors (AK) in the statistical package R (version 3.4.3; R Foundation for Statistical Computing) [47]. This explorative approach resulted in a new subgroup analysis in which we divided the control interventions into active (group discussion, mannequin-based simulation) and passive (lectures, reading assignments). Findings unsuitable for inclusion in a meta-analysis (eg, comparison of individual items in surveys) were presented using a narrative synthesis.

### Assessment of Risk of Bias

Two authors independently assessed the risk of bias using the Cochrane tool [40]. We considered the following domains: random sequence generation, allocation sequence concealment, blinding of participants or personnel, blinding to outcome assessment, completeness of outcome data, selective outcome reporting, and other sources of bias (eg, differences in baseline evaluation, volunteer bias, commercial grants). For cRCTs, we also assessed the risk of the following additional biases: recruitment bias, baseline imbalance, loss of clusters, incorrect analysis, and comparability with individually randomized trials. The publication bias in our review was difficult to investigate in a formal way because of high levels of heterogeneity which limit the interpretation possibilities of funnel plots.

### Summary of Findings Tables

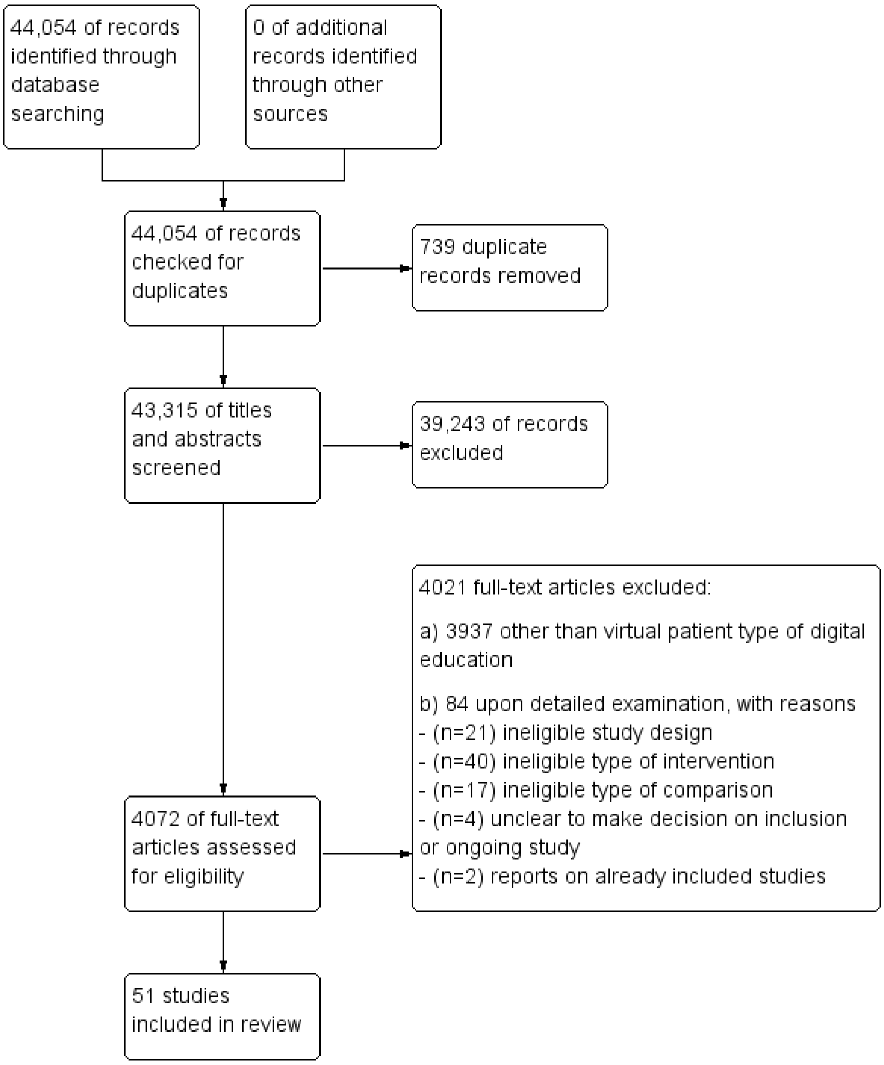
We prepared *summary of findings* tables to present results of the meta-analysis [40]. We presented the results for major comparisons of the review and for each of the major primary outcomes. We considered the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria to assess the quality of the evidence and downgraded the quality where appropriate [40].

## Results

### Included Studies

Our searches yielded a total of 44,054 citations, and 51 studies with 4696 participants were included (Figure 1). Overall, 2 reports described results already included in the review [48,49].

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




## Types of Studies

All included studies were published in peer-reviewed journals. All included studies had an RCT design, with the exception of 3 cRCTs [50-52].

## Types of Comparisons

A total of 25 studies compared virtual patients with traditional education [53-77], 11 compared a blend of virtual patients and traditional education with traditional education [50,52,78-86], 5 studies compared virtual patients with different forms of digital health education [51,87-90], and 10 studies compared different types of virtual patient interventions [91-100].

The traditional education control group involved a reading assignment in 6 studies [59,60,62,67-69]; 4 studies each involving a lecture [63,66,72,77], group assignment [53,65,71,73], and mannequin-based training [58,64,70,76]; and 1 each involving standardized patients [74] and ward-based education [75]. In 5 studies, the intervention was a mix of different forms of traditional education (eg, lecture, small group assignment, and mannequin-based training) [54-57,61].

The digital education control group was in 2 studies—a Web tutorial or course [88,90] and video recording [87,89]—and in 1 study, a mix of traditional lectures and Web materials including video clips [51] was used.

Studies comparing different types of virtual patients contrasted narrative with problem-solving structure of virtual patients [91]; virtual patients with and without usability enhancements [94]; different forms of feedback in virtual patients [95,96]; worked with unworked versions of virtual patients [97]; differences between self-determined and mandatory access to virtual patients [98]; virtual patients collections in which all the cases were presented at once to those automatically activated spaced in time [99]; effects of virtual patient solving with virtual patient construction exercises [100]; linear versus branched design of virtual patients [92]; and finally the addition of representation scaffolding (see glossary in Multimedia Appendix 1) to virtual patients [93].

Furthermore, 41 studies had 2 study arms (see the first table in Multimedia Appendix 3), 7 studies had 3 arms [62,88,91,95-98], and 3 studies had 4 arms [63,65,67].

## Types of Participants

In total, 41 studies involved preregistered professionals (see the first table in Multimedia Appendix 3), with 8 studies focused on postregistered participants [68,69,74,76,78,90,94,97]; 2

studies involved both pre- and postregistered participants [59,87].

In 37 out of 51 studies, participants were from the field of medicine. The studies from fields other than medicine were as follows: 6 studies in nursing [58,64,70,73,78,80]; 2 in pharmacy [53,92]; and 1 each in physical therapy [61], osteopathic medicine [84], and dentistry [83]. In addition, 3 studies involved interprofessional education [74,76,90].

A total of 44 out of 51 studies were conducted in high-income countries; 19 were from the United States (see the first table in Multimedia Appendix 3); 5 from Germany [65,82,93,98,99]; 3 each from Australia [52,70,91] and Sweden [83,85,87]; 2 each from Canada [59,60], the Netherlands [86,88], and the United Kingdom [66,77]; and 1 study was conducted each in Belgium and Switzerland [92], Denmark [100], France [54], Hong Kong [51], Japan [67], Poland [50], Singapore [64], and Slovenia [71]. From the 7 studies conducted in low-and-middle-income countries, 3 were from China [63,73,80], 2 from Colombia [55,56], and 1 each from the Republic of South Africa [94] and Iran [75].

In Multimedia Appendix 4 we present the technical characteristics of virtual patient systems, topics of educational content presented, applied instructional design methods, setting of use, information on the validity of outcome measurement, and applied educational theories in the included studies. Multimedia Appendix 5 summarizes the reasons for excluding studies following a review of their full-text versions.

## Effects of Interventions

### *Knowledge*

In total, 33 studies assessed outcomes of knowledge. In all studies, knowledge was measured using paper-based tests (see the second table in Multimedia Appendix 3). In 19 studies, the test consisted of multiple-choice questions (MCQs). Other knowledge test designs contained multiple-response questions [100], true/false questions [50], and key feature format questions [82]. In 4 studies, the participants had to formulate free-text answers [75,85,94,99]. In 4 studies [63,66,93,97], the knowledge tests comprised a mix of different formats. Li et al [63] used a combination of multiple-choice and short answer questions; Miedzybrodzka et al [66] used MCQs and modified essays; Harris et al [97] applied MCQs with confidence levels combined with script concordance testing questions; and Braun et al [93] used a test consisting of multiple-choice items, key feature problems, and problem-solving tasks. Secomb et al [70] measured cognitive growth using a survey requiring selection of the most significant items regarding learning environment preferences. In 3 studies [62,73,92], the nature of the knowledge test was unclear. In the case of MCQs in which the nature of items was unclear or mixed, we classified the outcome as knowledge instead of, for example, clinical reasoning skills, but the borderline between those was sometimes blurred. Meta-knowledge (eg, knowledge about the clinical reasoning process itself) was classified as knowledge outcomes following the framework by Kraiger [101].

The effects of interventions on knowledge outcomes are summarized in the second table in Multimedia Appendix 3.

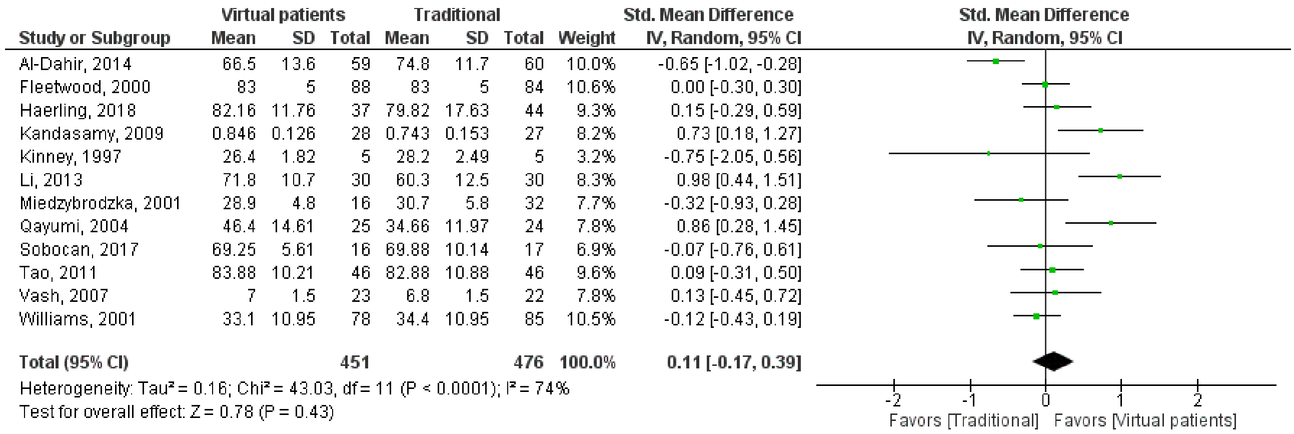
### Virtual Patient Versus Traditional Education

In 4 [60,63,67,72] of 18 studies comparing virtual patients with traditional education, the intervention resulted in more positive knowledge outcomes. In 2, the control group attended a lecture [63,72], whereas in the remaining 2 studies, students participated in a reading exercise [60,67]. In 1 study [53], the control intervention arm (Problem-based learning (PBL) small group discussion) had significantly better results than the virtual patient intervention (SMD=−0.65, 95% CI −1.02 to −0.28, $P$=.001). In the remaining 13 studies, the difference did not reach a statistically significant level (see the second table in Multimedia Appendix 3).

We excluded 2 studies [62,65] from our meta-analysis because of missing crucial outcome data. Jeimy et al [59] presented outcomes of a knowledge test compared item-by-item and the study was therefore excluded from the meta-analysis. We also excluded 1 study [72] owing to its outlier value of SMD=12.5 being most likely because of reporting error and excluded another study [70] as we regarded meta-knowledge as very different from the other types of core knowledge outcomes.

The pooled effect for knowledge outcomes (SMD=0.11, 95% CI −0.17 to 0.39, $I^2$=74%, n=927; Figure 2) suggests that virtual patient interventions are as efficient as traditional education.

**Figure 2.** Forest plot of virtual patient to traditional education comparison for knowledge outcomes. df: degrees of freedom; IV: interval variable; random: random effects model.

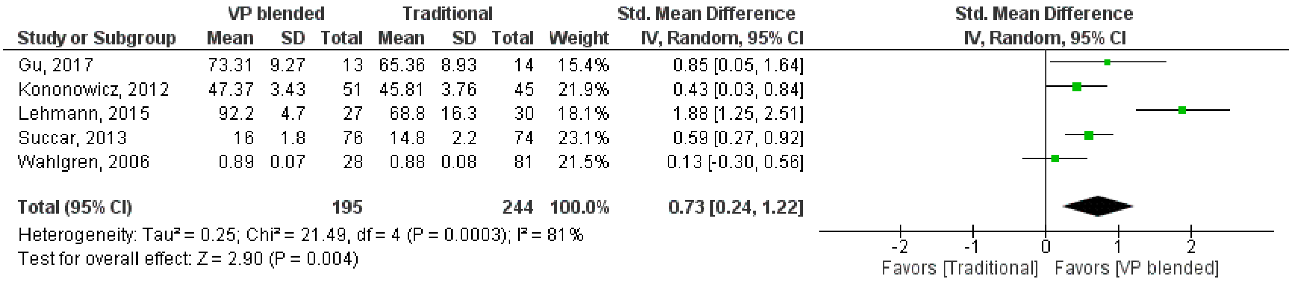


### Virtual Patient Blended Learning Versus Traditional Education

In 4 [50,52,80,82] of the 5 studies comparing virtual patients as a supplement with traditional education in the domain of knowledge, the group having the additional resource scored better than the control group. Only in 1 study [85] did the addition of virtual patients not lead to statistically significant difference in knowledge outcomes (*P*=.11).

The pooled effect for knowledge outcomes (SMD=0.73, 95% CI 0.24 to 1.22, $I^2$=81%, n=439; Figure 3) suggests moderate effects preferring the mix of virtual patients with traditional education over traditional education alone.

**Figure 3.** Forest plot of virtual patient blended learning to traditional education comparison for knowledge outcomes. df: degrees of freedom; IV: interval variable; random: random effects model; VP: virtual patients.




### Virtual Patient Versus Other Types of Digital Education

A total of 2 studies compared the difference in knowledge outcomes between virtual patients and digital health education interventions. Courteille et al [87] compared virtual patients with a video-recorded lecture, whereas Trudeau et al [90] compared with a static Web course. Neither of these comparisons showed significant differences in knowledge outcomes.

### Virtual Patient Design Comparison

In total, 8 studies focused on detecting the difference between variants of virtual patient design in the domain of knowledge. Only in 1 study by Friedman et al [96] were the differences at a statistically significant level. In this study, the pedagogic design of virtual patients was better than problem-solving and high-fidelity designs (*P*<.01). Comparing linear and branched virtual patients [92], scaffolded versus nonscaffolded [93], worked and unworked examples [97], virtual patient with usability extensions [94], self-determined versus mandatory integration [98], spaced versus nonspaced release of cases [99], and virtual patient solving versus virtual patient design exercises [100] resulted in no significant differences in knowledge outcomes.

## *Skills*

A total of 28 studies assessed skills outcomes (see the third table in Multimedia Appendix 3). Skills were assessed by performance on a mannequin in 9 studies [50,54,64,68,69,73,80,82,88], by performance on a live standardized patient in 8 studies [57,67,78,81,89,91,95,100], and performance on virtual patients [93] and real patients [83] in 1 study each. In 6 studies, outcomes were measured by a written assignment involving description of photographed clinical cases [63], radiographs [65], carrying out and structuring a mental state examination based on videotaped material [77], solving paper cases [74,86], and a modular paper-based test [75]. In 2 studies [55,56], outcomes were measured by a mix of paper cases and virtual patients. Kumta et al [51] combined computer-based assessment, objective structured clinical examination (OSCE), and clinical examination comprising patients in the ward into 1 score.

The effects of interventions on skills outcomes are summarized in the third table in Multimedia Appendix 3.

### Virtual Patient Versus Traditional Education

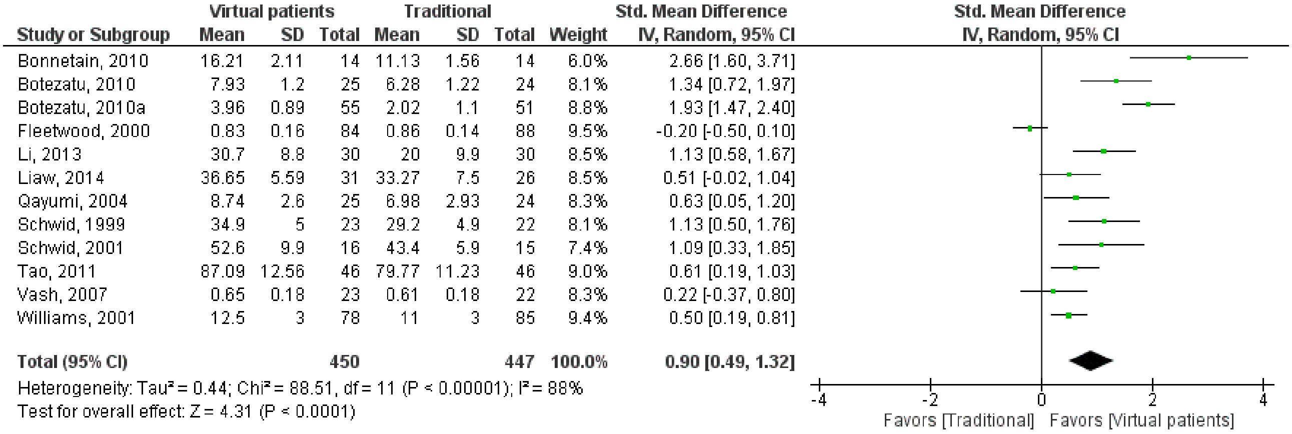
In 9 of 14 studies comparing virtual patients with traditional education, the intervention resulted in better skills outcomes (see the third table in Multimedia Appendix 3). The virtual patient intervention showed larger effects than lectures [63,77], reading exercises [67-69], group discussions [73], and activities comprising traditional methods, including lectures or hands-on training with mannequins [54-56].

Those skills which improved were clinical reasoning [55,56,63,77], procedural skills [54,67-69], and a mix of procedural and team skills [73].

We did not include in the meta-analysis 2 studies with incomplete reported data [65,74]. We also excluded skills outcomes from the study by Haerling [58] as these were available for a randomly selected subgroup only and from Wang et al [76] as they were measured for teams of students only and not individually.

The pooled effect on skills outcomes was (SMD=0.90, 95% CI 0.49 to 1.32, $I^2$=88%, n=897; Figure 4). Overall, this suggests that virtual patients have moderate to large positive effects in comparison with traditional education in the investigated types of skills.

**Figure 4.** Forest plot of virtual patient to traditional education comparison for skills outcomes. df: degrees of freedom; IV: interval variable; random: random effects model; VP: virtual patients.

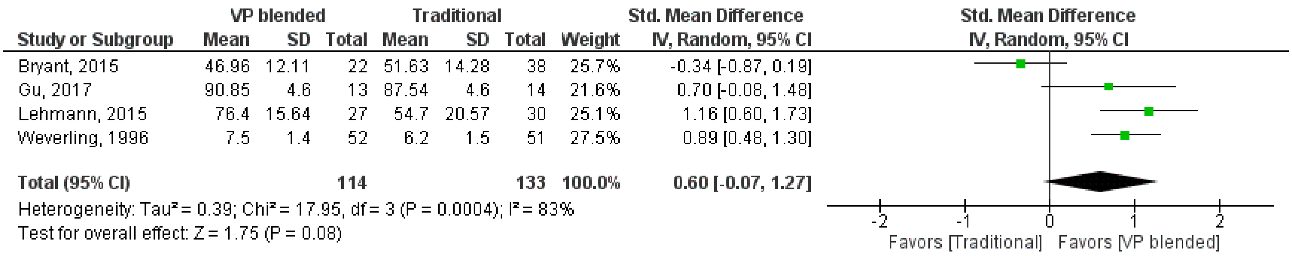



## Virtual Patient Blended Learning Versus Traditional Education

In 3 [82,83,86] out of 7 studies, the groups using virtual patients blended learning scored better than the control group in the skills domain. Lehmann et al [82] demonstrated significantly improved procedural skills (*P*<.001), whereas Weverling et al [86] reported on improved clinical reasoning skills (*P*<.001) and Schittek Janda et al [83] on communication skills (*P*<.01). Furthermore, 2 studies [78,80] involving nursing students showed no significant difference. The study by Bryant et al [78] evaluated communication skills (*P*=.38), whereas Gu et al [80] measured procedural skills (*P*>.05). We excluded 3 studies [50,81,83] from the meta-analysis because of insufficient data provided in the report or item-by-item comparison of a skills checklist.

The pooled effect for skills outcomes (SMD=0.60, 95% CI −0.07 to 1.27, $I^2$=83%, n=247; Figure 5) suggests that virtual patients blended with traditional education have moderate positive effects in comparison with traditional education alone.

**Figure 5.** Forest plot of virtual patient blended learning to traditional education comparison for skills outcomes. df: degrees of freedom; IV: interval variable; random: random effects model; VP: virtual patients.




## Virtual Patient Versus Other Types of Digital Education

Out of 3 studies comparing skills outcomes in virtual patients with other types of digital education studies, Kumta et al [51] showed a significant difference (*P*<.001). In this study, virtual patients were better than a range of different traditional teaching methods supplemented by Web content that included video clips, PowerPoint presentations, digital notes, and handouts. The target outcomes were clinical skills assessed by OSCE stations and examination of patients in the wards. In a study by Dankbaar et al [88], virtual patients were not significantly better in teaching procedural skill than an electronic module only (*P*>.05). Finally, in the study by Foster et al [89], virtual patients showed no significant difference when compared with video recordings in teaching communication skills (*P*>.05).

## Virtual Patient Design Comparison

From the 4 studies that compared the influence of different virtual patient designs on skills outcomes, Foster et al [95] showed that virtual patients with emphatic feedback were significantly better in training communication skills than those virtual patients without feedback (*P*<.03). In the study by Bearman et al [91], narrative virtual patients were significantly better than problem-solving virtual patients in conveying communication skills (*P*=.03). In a study by Braun et al [93], the addition of representational scaffolding to a virtual patient intervention significantly improved diagnostic efficiency (*P*=.045). Finally, in the study by Tolsgaard et al [100], there was no significant difference in integrated clinical performance when students constructed or solved virtual patients (*P*=.54).

## Attitudes

A total of 11 studies reported attitudinal outcomes (see the third table in Multimedia Appendix 1). The attitudes related to confidence, preparedness, comfort, self-efficacy, and perceived ability in topics such as history taking and clinical breast examination [79], diagnostic and management abilities [59], contrast reaction management and teamwork [76], ethical, legal, and communication issues [57,81], opioid therapy [90], cultural competence [84], procedural knowledge in pediatric basic life support [82], performing pharmacy triage [92], caring for distress disorders patients [74], and anxiety [77].

The effects of interventions on attitudinal outcomes are presented in the fourth table in Multimedia Appendix 3. Furthermore, 3 studies presented pooled scores on students‘ self-assessment. In the study by Lehmann et al [82], students felt more confident in their knowledge and skills on performing pediatric basic life support with additional access to virtual patients that supplemented their traditional course ($P$<.001). There were no significant differences in the remaining 2 studies focusing on communication-related self-efficacy [81] and attitudes related to opioid therapy [90].

In the study by Williams et al [77], more items related to self-assessment of competences (in dealing with ethical aspects and managing anxiety) were scored lower in the virtual patient group than in the traditional education groups. There were no differences in analyzed items related to attitudes in 4 studies [57,59,74,79]. In the study by Smith et al [84], the results regarding attitudes toward clinical cultural competence were presented separately for bilingual and English-speaking students, which makes it difficult to aggregate not knowing the number of bilingual students in each study group. However, the descriptive conclusion of the authors was that general cultural competence measures were the same for the virtual patient and control group. In 2 studies [76,92], the results were compared item-by-item and only within the groups (pre/posttest), not between the study groups.

### Satisfaction

In total, 17 studies measured satisfaction resulting from an intervention (see the fifth table in Multimedia Appendix 3). All outcomes in this category were measured by satisfaction questionnaires. Different facets of satisfaction were measured, which we classified in the following 5 dimensions: general impression (global score or willingness to recommend), comfort in use (learning style preference, engagement or motivation, positive climate or safety, and enjoyment or pleasure), integration in curriculum (time constraints, relevance, and level of difficulty), academic factors (feedback quality, structure, and clarity), and satisfaction with technical features (usability and information technology readiness).

In 4 out of 17 studies evaluating the satisfaction of students receiving a virtual patient intervention, the result was presented as 1 aggregated score of several items. Furthermore, 3 of those studies compared different design variants of virtual patients. In the study by Friedman et al [96], the pedagogic format (menus, guided) resulted in higher satisfaction scores than the high-fidelity (free text, unguided) format ($P$<.01). There was no statistically significant difference between the virtual patients with and without usability enhancements [94] ($P$=.13) and solving versus constructing virtual patients [100] ($P$=.46). One study [58] presented comparison of virtual patients with mannequin-based training using a single score for student satisfaction and self-confidence in learning, showing no difference between the simulation modalities ($P$=.11).

In the remaining 13 out 17 studies, the survey responses were compared item-by-item. In 4 studies, the majority of the items indicated preference for the virtual patient intervention, in comparison with lecture [63], reading assignment [67], video-based learning [89], and Web tutorial [88]. In 7 studies, most items were indifferent between the groups [53,59,62,65,66,74,92]. In 1 study [76], most items (5 out of 6) in a satisfaction survey were better rated in the mannequin-based training than in the virtual patient group.

## Secondary Outcomes

One study had cost-effectiveness as an outcome [58]. In 9 studies, statements were made regarding the cost of the intervention—either monetary or in development time [53,60,62,64-66,79,95]. Only 1 study provided numerical data on both types of intervention [95]. The comparison was qualitative in 3 studies [64,65,78]. In 5 studies, estimations of costs were made for the virtual patient group without contrasting it with the cost of the control intervention [53,60,62,66,79]. None of the included studies had patient outcomes or adverse effects as the main outcome measure. Even though none of the studies reported direct patient outcomes, in 2 studies, the participants were observed by raters while performing tasks on real patients as an outcome assessment [51,83]. In the study by Kumta et al [51], the score was included in more complex assessment (including MCQ tests and OSCE examination) and the patient-related outcome was not explicitly reported. In the study by Schittek Janda et al [83], first year students of dentistry were asked to perform history taking with real patients and were rated by the instructor. The patients’ perspective was, however, not considered. Even though none of the studies had adverse effects as the major outcome, 6 studies [53,55,67,70,84,88] reported findings related to noticed unexpected effects of the intervention.

### *Cost*

Haerling [58] showed a better cost-utility ratio of US $1.08 for virtual patients versus US $3.62 for the mannequin-based training. Foster et al [95] compared the cost of human-provided (Mechanical Turk) feedback with backstory video feedback; the cost of human answers was US $0.05 per question assisted, whereas videos required 4 hours of development time and the license cost of a video game (Sims 3 by Electronic Arts). This does not provide a direct answer to the question of which method was more cost-efficient as it depends on the number of participants and time of use. It is also important to notice that the human-generated feedback in virtual patients showed positive effects on the communication skills outcomes, whereas the backstory video did not. Bryant et al [78] estimated, but without providing numerical evidence, that the cost of a virtual patient was similar to that of a course text that was eliminated by the new intervention. Liaw et al [64], without providing concrete numbers, noticed that despite “initial startup costs for developing the virtual patient simulation, its implementation was less resource intensive than the mannequin-based simulation.” The cost savings were because of reduced instructor time, use of expensive equipment, or simulation facilities. Maleck et al [65] saw cost savings in the virtual patient group because of spared radiograph printouts. The cost of the virtual patient intervention was expressed in hours of work; in 2 cases, the cost was 12 to 15 hours per virtual patient [53,60]; in 1 case it was 15 to 30 hours [62] and 100 hours in another [66]. The cost expressed in amounts of money was estimated at US $500 for content development and technical implementation [62] and

US $4800 for a total clerkship restructuring, including adding virtual patients [60]. It is worth noticing that in both cases the virtual patients were developed by students. Deladisma et al [79] used in their study a virtual patient system that involved a speech recognition engine, tracked user's body movements, and projected a life-sized avatar on the wall. The cost of the technology used in the pilot study (including 2 networked personal computers, 1 data projector, and 2 Web cameras) was estimated in 2006 to be less than US $7000 [102].

### *Patient Outcomes*

In the study by Schittek Janda et al [83], an experienced clinician rated the professional behavior (language precision, order of question, and empathy) of first year students' of dentistry toward real patients as significantly higher ($P$<.01) in the group having access to a supplementary virtual patient case than in the group that underwent standard instruction.

### *Adverse Effects*

Dankbaar et al [88] hypothesize based on their study results that high-fidelity virtual patients may increase motivation, but at the same time be more distracting for novice students and by that impede learning. Authors of 2 studies [70,84] observe that the language of virtual patients might be a significant factor showing greater effects on nonnative English speaking and bilingual learners than in native English speakers. In the study by Qayumi et al [67], it is observed that lower-achieving students benefit more from virtual patients than high performers. In the study by Botezatu et al [55], students knowing about the possibility of being assessed by virtual patients opposed being tested with paper cases. In the study by Al-Dahir et al [53], it is observed that analysis of individual learner traces in the virtual patient system negates benefits of social learning.

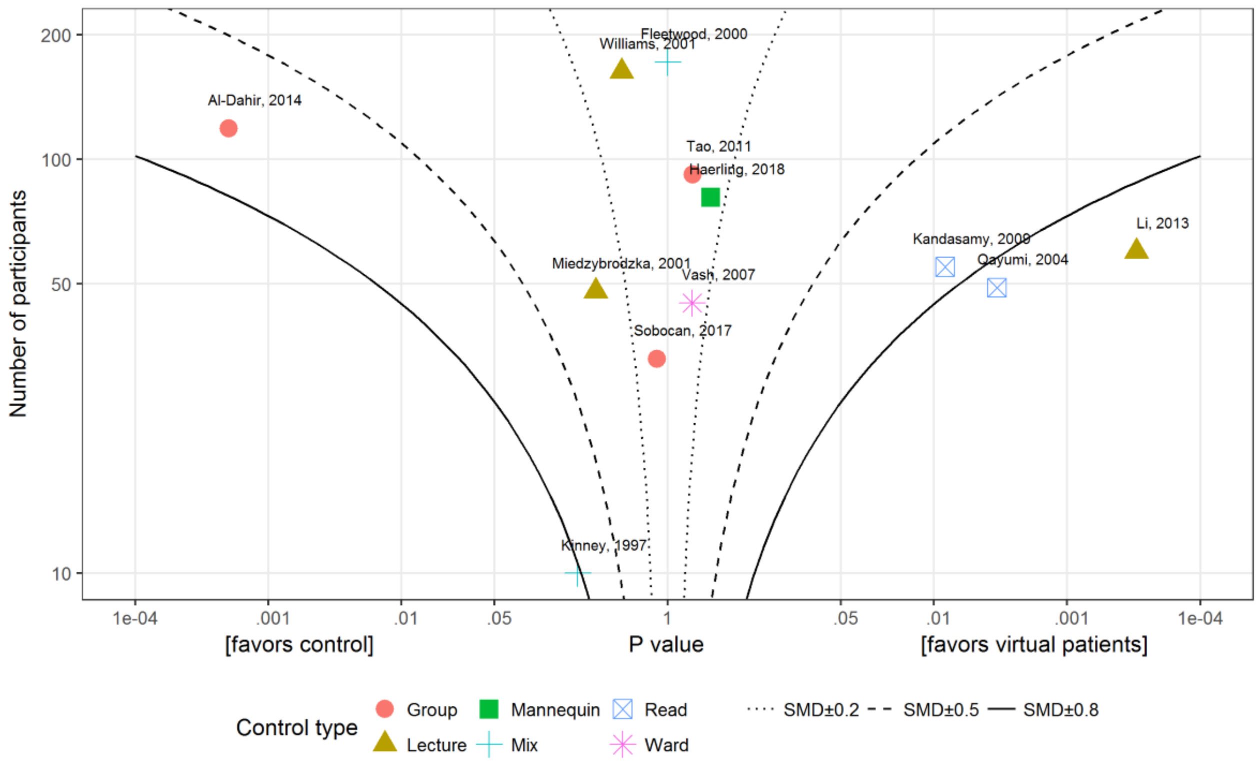
## Subgroup Analysis

None of the initially planned subgroup analyses explained the heterogeneity of the results.

Among many analyzed aspects, we looked into differences regarding the efficiency of learning with virtual patients between the health professions disciplines. Most of the located studies involved students of medicine as participants. For instance, when comparing virtual patients with traditional education in the domain of skills, out of the 12 outcomes included for subgroup analyses, only 2 were from other health profession disciplines than medicine (ie, studies from nursing [64,73]). When analyzing knowledge outcomes out of the 12 included studies, 4 were nonmedical but represented 3 very different disciplines, nursing [58,73], pharmacy [53], and physiotherapy [61]. The conducted subgroup analyses showed no significant differences between the subgroups and high heterogeneity.

While analyzing aspects of instructional design implemented in the virtual patient scenarios, we were able to locate a very balanced number of studies implementing the narrative and problem-solving designs [91] in the domain of knowledge outcomes (6 studies in each branch). Yet, the pooled results showed no difference (narrative: SMD=0.12, 95% CI −0.41 to 0.64, $I^2$=85%, n=525 versus problem solving: SMD=0.11, 95% CI −0.17 to 0.38, $I^2$=51%, n=520; subgroup differences $P$=.97). Interestingly, when looking into the domain of skills outcomes, all studies had either the problem-solving or unclear design (in 2 cases). This might be an indication that narrative (linear, branched) virtual patients are seen as being better suited for knowledge outcomes rather than skills.

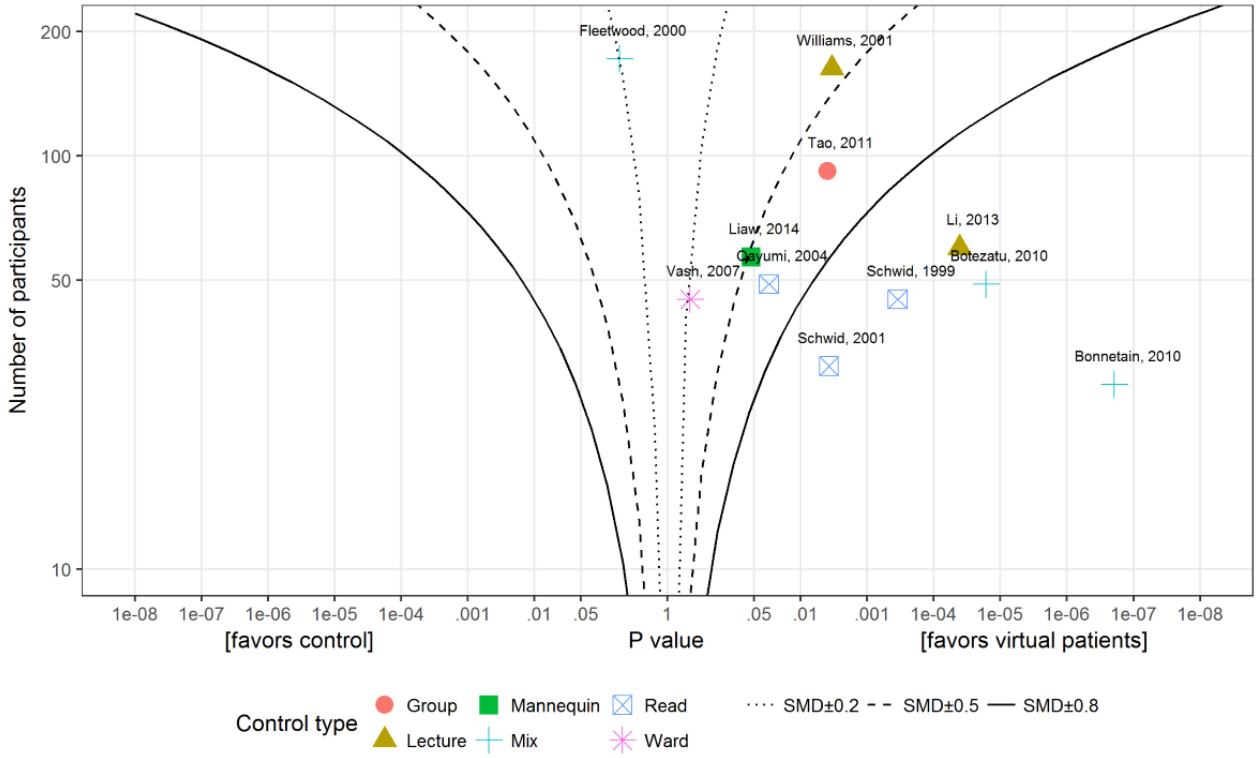
**Figure 6.** Albatross plot for studies comparing virtual patient with traditional education for knowledge outcomes. SMD: standardized mean difference.



Finally, we were unable to see any pattern in efficiency when analyzing the timing of feedback as being either during activity or post activity. However, in almost half of the studies, we were unable to decide, based on the description of the intervention, which model of feedback was implemented or whether the study had a mixed (during/post activity) mode of providing feedback.

To further explore the reasons for heterogeneity, we visualized the outcomes in the form of albatross plots of the knowledge and skills outcomes for virtual patients to traditional education comparisons. Figure 6 presents an albatross plot for knowledge and Figure 7 for skills outcomes. Comparisons of virtual patients to passive forms of learning (reading exercises and lectures) tended to display large positive effect sizes, whereas those comparing virtual patients to active learning (group discussion or mannequin-based learning) show small effects or even negative effects (left hand side in the Figures 6 and 7 and Multimedia Appendix 6).

**Figure 7.** Albatross plot for studies comparing virtual patient with traditional education for skills outcome. SMD: standardized mean difference.



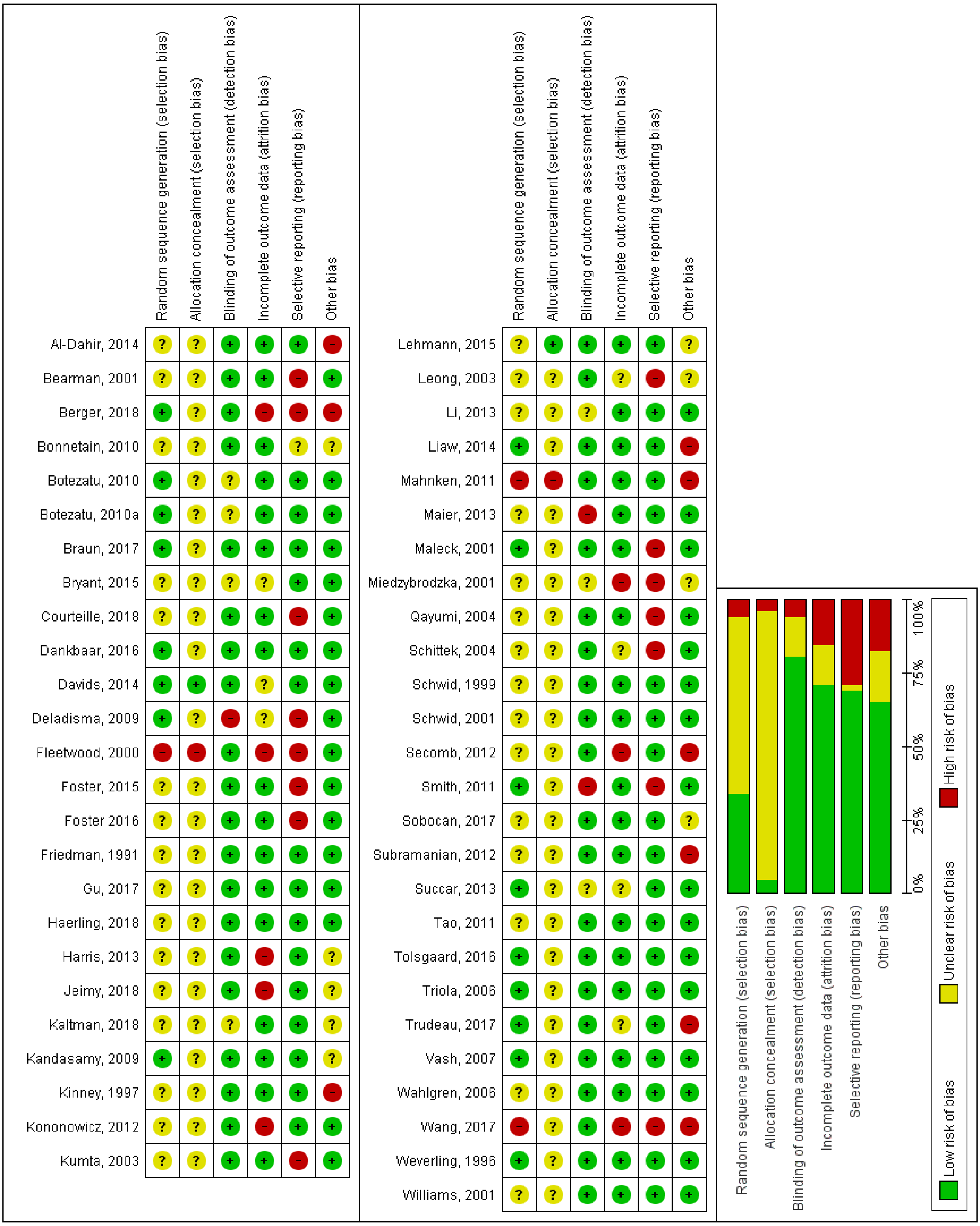

## Risk of Bias

Following the Cochrane methodology [40], we have assessed the risk of bias in all included studies. The results of the analysis are summarized in Figure 8.

Overall, we do not consider allocation bias as a significant issue in the review as most of the studies either described an adequate randomization method (17 of 51 studies) or even when the description was unclear (31 of 51), it was judged unlikely that the randomization was seriously flawed. Performance bias in comparisons with traditional education is an issue but at the same time is impossible to avoid in this type of research. The blinding of participants in virtual patient design comparisons is possible, but those studies are still relatively uncommon (n=10). The risk of assessor bias was avoided in many studies by using automated or formalized assessment instruments. Consequently, we assessed the risk as low in 42 of 51 studies. However, it is often unclear whether the instruments (eg, MCQ tests, assessment rubrics) were properly validated. We felt that in the majority of studies, attrition bias was within acceptable levels (low risk in 36 of 51 studies). This does not exclude volunteer bias, which is likely to be common, but its influence is difficult to estimate. As there is little tradition of publishing protocols in medical education research, it was problematic to assess selective reporting bias, but we judged the risk as low in 35 out of 51 studies. We were unable to reliably assess publication bias considering the high heterogeneity of studies. None of the cRCT studies considered in the statistical analysis had corrections for clustering, but we have decreased the number of participants in those studies using a method from the Cochrane Handbook to compensate for that. We present more details of the risk of bias analysis in Multimedia Appendix 7.

We rated down the quality of evidence for knowledge and skills outcomes in virtual patients to traditional education comparison because of the high heterogeneity of included studies and limitations in study design (lack of participant blinding, nonvalidated instruments, and potential volunteer bias). For attitudinal and satisfaction outcomes and for other types of comparisons, we additionally rated down the quality as the outcomes were presented as independent items in questionnaires that were not amenable to statistical analysis or the analyses contained just a handful of studies and the CIs were wide. Summary of findings table (GRADE) are presented in Multimedia Appendix 8.

**Figure 8.** Risk of bias summary (+ low risk of bias; - high risk of bias, ? unclear risk of bias).




## Discussion

### Principal Findings

The aim of this review was to evaluate the effectiveness of virtual patients in comparison with other existing educational methods.

There is low quality evidence that virtual patients are at least as effective as traditional education for knowledge outcome and more effective for skills outcomes. On the basis of the visual analysis of albatross plots, we may hypothesize that replacing passive forms of traditional education with virtual patients brings more benefit than replacing active learning methods. We collected positive evidence of effectiveness from both high-income and low-and-middle-income countries demonstrating the global applicability of virtual patients. Students were generally satisfied with the use of virtual patients, but we also located studies in our review where the use of virtual patients was connected with diminished confidence.

The strength of our systematic review is the broad perspective which shows the landscape of RCTs in the domain of virtual patients. Our systematic review updates the evidence on virtual patient effectiveness, which was last summarized in a meta-analysis almost a decade ago.

### Limitations

The limitation of our work is that the wide scope of the review does not allow nuances in the studies to be explored in detail. We were unable to make a firm assessment of publication bias. The high heterogeneity of the results leads to the conclusion that without further consideration of needs and implementation details, we cannot expect that the introduction of virtual patients will always lead to detectable positive outcomes. Evidence to determine the effective factors is sparse and represented by only 10 studies in our review, with very diverse research questions.

Our review is limited by the decision to exclude crossover design studies. However, this has been discussed in detail in the potential biases in the review process section in Multimedia Appendix 7. We excluded studies published before 1991 as we consider the technology available before the World Wide Web to be materially different from that currently available. Finally, we are limited by the sparse description of the interventions in some of the papers, which occasionally might have led to misclassification of the studies.

### Comparison With Prior Work

Extending the results of the meta-analysis by Cook et al [38] and in agreement with the one by Consorti et al [39], our review shows that virtual patients have an overall positive pooled effect when compared with some other types of traditional educational methods. Our observations regarding the influence of the type of outcome (knowledge/skills) and comparison (active/passive traditional learning) supplement the evidence in the previous reviews [38,39], which included studies until 2010. This time point divides the evidence collected in 2 parts: (1) time already covered by previous reviews (1991-2010) and (2) time not included in the previous reviews (2011-2018). It is interesting to note that, though the former timeframe spans over 20 years compared with 8 years in the latter, more studies were included from the latter period, 22 studies (until 2010) versus 29 studies (after 2010). This demonstrates increased interest in virtual patients and medical education in general. The research community around digital health education has long been criticized for publishing media-comparative studies [103,104]. Media-comparative research aims to make comparisons between different media formats such as paper, face-to-face, and digital education [104]. Both Friedman and Cook argue [103,104] that the limitations of this type of comparison boils down to the inability to produce an adequate control group as interventions are bound to be influenced by too many confounding factors to be generalizable. Even though there are still many media-comparative studies, the number of studies comparing different forms of digital education seem to increase: 3/22 (14%) until 2010 versus 11/29 (38%) after 2010. The number of studies in which students worked from home as an intervention has also increased; before 2011 there was just 1/16 (6%; in 6 studies it was unclear), whereas after 2011, it was 11/22 (50%; in 7 studies it was unclear) studies. However, this potentially raises concerns about how controlled the interventions and measures were, and thus the validity of the conclusions.

Our observation that virtual patient simulations predominantly effect skills rather than knowledge outcomes can be interpreted as an indication that for lower levels of Bloom's taxonomy [105], (*remember, understands*) there is little added value of introducing virtual patients when compared with traditional methods of education. Virtual patients can have greater impact when applied where knowledge is combined with skills and applied in problem solving, and when direct patient contact is not yet possible. We found little evidence to support the use of virtual patients at higher levels of the taxonomy. We also warn against using our result in justifying diminished hours of bedside teaching as this was investigated in just 1 study [75] and did not show positive outcomes. Consequently, virtual patients can be said to be a modality for learning in which learners actively use and train their clinical reasoning and critical thinking abilities before bedside learning, as was previously suggested in their critical literature review by Cook and Triola [22].

The perceptions of students toward studying with virtual patients are generally positive. However, some exceptions can be noted. In 1 study [77], students were less confident in their skills when compared with facilitated group discussion and lecture. This is in contrast with no observable differences or even better performance in the virtual patient group when considering the objective outcomes in those studies. This could be explained by disbelief in the effectiveness of the new computer-based methods of learning or anxiety of losing direct patient contact.

The results of our subgroup analysis, though inconsistent, encourage the introduction of more active forms of education. Yet, we note that the range from active to passive learning forms a continuum, and the decision on how to classify each intervention is hampered by sparse descriptions in the reports. Nevertheless, questioning the utility of passive learning is not a new finding and is observed elsewhere, for instance, in the literature on the flipped-classroom learning approach [106]. As the effects of comparing virtual patients with other forms of active learning were small and we could not detect any other variables explaining the heterogeneity, it seems reasonable to individually consider other factors such as cost of use, time flexibility, personnel shortage, and availability in different settings (eg, students' homes or locations remote from academic centers) when determining which methods to use.

The need for more guidance within virtual patient simulations is apparent in studies differing by instructional methods where narrative virtual patient design was better than more autonomous problem-oriented designs [91]. Feedback given by humans at distance in a virtual patient system was better than an animated backstory in increasing empathy [95], whereas more active constructing virtual patients with more time on a task but no feedback had no more positive result on the outcomes than learning from a virtual patient scenario [100]. This reminds us that presenting realistic patient scenarios with a great degree of freedom cannot be an excuse for neglecting guidance in relation to learning objectives [107,108].

## Outlook

We join the plea of Friedman [103] and Cook [104] to abandon media-comparative research as it is difficult to interpret and we instead encourage greater focus on exploring the utility of different design variants of virtual patient simulations. The current knowledge on the influence of these factors is sparse. A carefully planned study backed up in sound educational theory should provide many valuable research opportunities. However, sufficiently powered samples are needed, as the effects are likely to be small. The second consideration pertains to the need to use previously validated measurement tools that are well-aligned with the learning objectives. Comparisons of outcomes in tools on an item-by-item basis is methodologically questionable and makes the aggregations of results difficult in systematic reviews. We also call for more studies in other health professions disciplines than medicine, as our subgroup analysis showed that evidence of virtual patient effectiveness in such programs as nursing, physiotherapy, or pharmacy is underrepresented. Investigations into patient outcomes and cost-effectiveness of virtual patients are not yet explored directly and form a key avenue for future efforts.

## Conclusions

Low to modest and mixed evidence suggests that when compared with traditional education, virtual patients can more effectively improve skills, and at least as effectively improve knowledge outcomes as traditional education. Education with virtual patients provides an active form of learning that is beneficial for clinical reasoning skills. Implementations vary and are likely to be broad across pre- and postregistration education, although current studies do not provide clear guidance on when to use virtual patients. We recommend further research be focused on exploring the utility of different design variants of virtual patients.

## Acknowledgments

The authors wish to acknowledge the help of Jayne Alfredsson, Carina Georg, Anneliese Lilienthal, Italo Masiello, and Monika Semwal at the stage of protocol writing. The authors would like to thank Jim Campbell for his contribution to coordination and conceptual formation of the Digital Health Education Collaboration. The authors would also like to thank Carl Gornitzki, GunBrit Knutssön, and Klas Moberg from the University Library, Karolinska Institute, for developing the search strategy and Ram Chandra Bajpai from Nanyang Technological University for statistical consulting and Pawel Posadzki for his supportive feedback on the manuscript.

At the time of the study, SE was affiliated with the Karolinska Institutet and Linköping University, and N Saxena was affiliated with Health Services and Outcomes Research, National Healthcare Group, Singapore, Singapore.

## Authors' Contributions

JC conceived the idea for the review. AK drafted the protocol with substantial contributions from LW, SE, NS, DD, NSA, LTC, and NZ. The Digital Health Education Collaboration developed the search strategy, obtained copies of studies, and screened the studies. JC and JCD contributed to the coordination and conceptual formation of the Digital Health Education Collaboration. AK, LW, NS, and DD extracted data from studies and conducted the risk of bias assessment. LW contacted the authors in case of missing data. AK carried out the analysis of collected data. LTC provided methodological guidance. AK, LW, SE, NS, DD, LTC, and NZ interpreted the analysis and contributed to the discussion. AK drafted the final review with substantial contributions from all authors. All authors revised and approved the final version of review.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Glossary.

[DOC File, 76KB - jmir_v21i7e14676_app1.doc ]

## Multimedia Appendix 2

MEDLINE (Ovid) search strategy.

[DOC File, 47KB - jmir_v21i7e14676_app2.doc ]

## Multimedia Appendix 3

Summary of included studies.

[DOC File, 291KB - jmir_v21i7e14676_app3.doc ]

## Multimedia Appendix 4

Summary of technical and educational features of included studies.

[DOC File, 458KB - jmir_v21i7e14676_app4.doc ]

## Multimedia Appendix 5

Summary of excluded studies.

[DOC File, 60KB - jmir_v21i7e14676_app5.doc ]

## Multimedia Appendix 6

Subgroup analysis details.

[DOC File, 66KB - jmir_v21i7e14676_app6.doc ]

## Multimedia Appendix 7

Quality of the evidence.

[DOC File, 89KB - jmir_v21i7e14676_app7.doc ]

## Multimedia Appendix 8

Summary of finding tables.

[DOC File, 86KB - jmir_v21i7e14676_app8.doc ]

## References


1. Ramani S, Leinster S. AMEE Guide no. 34: teaching in the clinical environment. Med Teach 2008 Jan;30(4):347-364. [doi: 10.1080/01421590802061613] [Medline: 18569655]
2. Moalem J, Salzman P, Ruan DT, Cherr GS, Freiburg CB, Farkas RL, et al. Should all duty hours be the same? Results of a national survey of surgical trainees. J Am Coll Surg 2009 Jul;209(1):47-54, 54.e1-2. [doi: 10.1016/j.jamcollsurg.2009.02.053] [Medline: 19651062]
3. Dev P, Schleyer T. Computers in health care education. In: Shortliffe E, Cimino J, editors. Biomedical Informatics. London: Springer; 2014:675-693.
4. 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. Lancet 2010 Dec 4;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
5. Crisp N, Gawanas B, Sharp I, Task Force for Scaling Up EducationTraining for Health Workers. Training the health workforce: scaling up, saving lives. Lancet 2008 Feb 23;371(9613):689-691. [doi: 10.1016/S0140-6736(08)60309-8] [Medline: 18295028]
6. 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]
7. Bajpai S, Semwal M, Bajpai R, Car J, Ho AH. Health professions' digital education: review of learning theories in randomized controlled trials by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 12;21(3):e12912 [FREE Full text] [doi: 10.2196/12912] [Medline: 30860483]
8. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GC, Tudor Car L. Mobile digital education for health professions: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937 [FREE Full text] [doi: 10.2196/12937] [Medline: 30747711]
9. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
10. George PP, Zhabenko O, Kyaw BM, Antoniou P, Posadzki P, Saxena N, et al. Online digital education for postregistration training of medical doctors: systematic review by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 25;21(2):e13269 [FREE Full text] [doi: 10.2196/13269] [Medline: 30801252]
11. Huang Z, Semwal M, Lee SY, Tee M, 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 2019 Feb 21;21(2):e12997 [FREE Full text] [doi: 10.2196/12997] [Medline: 30789348]

12. Kyaw BM, Posadzki P, Dunleavy G, Semwal M, Divakar U, Hervatis V, et al. Offline digital education for medical students: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 25;21(3):e13165 [FREE Full text] [doi: 10.2196/13165] [Medline: 30907731]
13. 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]
14. Lall P, Rees R, Law GC, Dunleavy G, Cotic 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 Feb 28;21(2):e12895 [FREE Full text] [doi: 10.2196/12895] [Medline: 30816847]
15. Martinengo L, Yeo NJY, Tang ZQ, Markandran KD, Kyaw BM, Tudor Car L. Digital education for the management of chronic wounds in health care professionals: protocol for a systematic review by the Digital Health Education Collaboration. JMIR Res Protoc 2019 Mar 25;8(3):e12488 [FREE Full text] [doi: 10.2196/12488] [Medline: 30907743]
16. Posadzki P, Bala MM, Kyaw BM, Semwal M, Divakar U, Koperny M, et al. Offline digital education for postregistration health professions: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Apr 24;21(4):e12968 [FREE Full text] [doi: 10.2196/12968] [Medline: 31017584]
17. Semwal M, Whiting P, Bajpai R, Bajpai S, Kyaw BM, Tudor Car L. Digital education for health professions on smoking cessation management: systematic review by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 4;21(3):e13000 [FREE Full text] [doi: 10.2196/13000] [Medline: 30829576]
18. Tudor Car L, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, et al. Digital problem-based learning in health professions: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 28;21(2):e12945 [FREE Full text] [doi: 10.2196/12945] [Medline: 30816846]
19. Wahabi HA, Esmaeil SA, Bahkali KH, Titi MA, Amer YS, Fayed AA, et al. Medical doctors' offline computer-assisted digital education: systematic review by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 1;21(3):e12998 [FREE Full text] [doi: 10.2196/12998] [Medline: 30821689]
20. Ellaway R, Candler C, Greene P, Smothers V. MedBiquitous. An Architectural Model for MedBiquitous Virtual Patients URL: http://tinyurl.com/jpewpbt [accessed 2019-05-10] [WebCite Cache ID 6qkBNcQGi]
21. Kononowicz AA, Zary N, Edelbring S, Corral J, Hege I. Virtual patients - what are we talking about? A framework to classify the meanings of the term in healthcare education. BMC Med Educ 2015;15:11 [FREE Full text] [doi: 10.1186/s12909-015-0296-3] [Medline: 25638167]
22. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
23. Posel N, McGee JB, Fleiszer DM. Twelve tips to support the development of clinical reasoning skills using virtual patient cases. Med Teach 2014 Dec 19:1-6. [doi: 10.3109/0142159X.2014.993951] [Medline: 25523009]
24. Berman NB, Durning SJ, Fischer MR, Huwendiek S, Triola MM. The role for virtual patients in the future of medical education. Acad Med 2016 Sep;91(9):1217-1222. [doi: 10.1097/ACM.0000000000001146] [Medline: 26959224]
25. Kolb DA. Experiential Learning: Experience As The Source Of Learning And Development. Upper Saddle River, New Jersey: Prentice Hall; 1984.
26. Yardley S, Teunissen PW, Dornan T. Experiential learning: AMEE Guide No. 63. Med Teach 2012;34(2):e102-e115. [doi: 10.3109/0142159X.2012.650741] [Medline: 22289008]
27. Edelbring S, Dastmalchi M, Hult H, Lundberg IE, Dahlgren LO. Experiencing virtual patients in clinical learning: a phenomenological study. Adv Health Sci Educ Theory Pract 2011 Aug;16(3):331-345. [doi: 10.1007/s10459-010-9265-0] [Medline: 21656337]
28. Norman G. Research in clinical reasoning: past history and current trends. Med Educ 2005 Apr;39(4):418-427. [doi: 10.1111/j.1365-2929.2005.02127.x] [Medline: 15813765]
29. Berman N, Fall LH, Smith S, Levine DA, Maloney CG, Potts M, et al. Integration strategies for using virtual patients in clinical clerkships. Acad Med 2009 Jul;84(7):942-949. [doi: 10.1097/ACM.0b013e3181a8c668] [Medline: 19550193]
30. Kenny NP, Beagan BL. The patient as text: a challenge for problem-based learning. Med Educ 2004 Oct;38(10):1071-1079. [doi: 10.1111/j.1365-2929.2004.01956.x] [Medline: 15461652]
31. Button D, Harrington A, Belan I. E-learning & information communication technology (ICT) in nursing education: a review of the literature. Nurse Educ Today 2014 Oct;34(10):1311-1323. [doi: 10.1016/j.nedt.2013.05.002] [Medline: 23786869]
32. Watson JA, Pecchioni LL. Digital natives and digital media in the college classroom: assignment design and impacts on student learning. Educ Media Int 2011 Dec;48(4):307-320. [doi: 10.1080/09523987.2011.632278]
33. Schifferdecker KE, Berman NB, Fall LH, Fischer MR. Adoption of computer-assisted learning in medical education: the educators' perspective. Med Educ 2012 Nov;46(11):1063-1073. [doi: 10.1111/j.1365-2923.2012.04350.x] [Medline: 23078683]
34. Baumann-Birkbeck L, Florentina F, Karatas O, Sun J, Tang T, Thaung V, et al. Appraising the role of the virtual patient for therapeutics health education. Curr Pharm Teach Learn 2017 Dec;9(5):934-944. [doi: 10.1016/j.cptl.2017.05.012] [Medline: 29233327]

35. Cendan J, Lok B. The use of virtual patients in medical school curricula. Adv Physiol Educ 2012 Mar;36(1):48-53 [FREE Full text] [doi: 10.1152/advan.00054.2011] [Medline: 22383412]
36. Poulton T, Balasubramaniam C. Virtual patients: a year of change. Med Teach 2011 Jan;33(11):933-937. [doi: 10.3109/0142159X.2011.613501] [Medline: 22022903]
37. Saleh N. The value of virtual patients in medical education. Ann Behav Sci Med Educ 2010 Oct 16;16(2):29-31. [doi: 10.1007/BF03355129]
38. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]
39. Consorti F, Mancuso R, Nocioni M, Piccolo A. Efficacy of virtual patients in medical education: a meta-analysis of randomized studies. Comput Educ 2012 Nov;59(3):1001-1008. [doi: 10.1016/j.compedu.2012.04.017]
40. Higgins J, Green S. Cochrane Handbook For Systematic Reviews Of Interventions. Chichester, England: John Wiley & Sons; 2008.
41. Kononowicz AA, Woodham L, Georg C, Edelbring S, Stathakarou N, Davies D, et al. Virtual patient simulations for health professional education. Cochrane Database Syst Rev 2016 May 19(5):CD012194. [doi: 10.1002/14651858.CD012194]
42. 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]
43. EndNote X7. Philadelphia, PA: Thomson Reuters; 2016. URL: https://endnote.com/ [accessed 2019-05-10]
44. 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]
45. Cochrane Community. Copenhagen, Denmark: The Nordic Cochrane Centre; 2014. Review Manager (RevMan) Version 5.3 URL: https://community.cochrane.org/help/tools-and-software/revman-5 [accessed 2019-05-10]
46. Harrison S, Jones HE, Martin RM, Lewis SJ, Higgins JP. 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]
47. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Core Team; 2017. URL: https://www.r-project.org/ [accessed 2019-05-10]
48. Kurihara Y, Kuramoto S, Matsuura K, Miki Y, Oda K, Seo H, et al. Academic performance and comparative effectiveness of computer- and textbook-based self-instruction. Stud Health Technol Inform 2004;107(Pt 2):894-897. [Medline: 15360941]
49. Succar T, Grigg J. A new vision for teaching ophthalmology in the medical curriculum: The Virtual Ophthalmology clinic. 2010 Presented at: ASCILITE Conference; December 5-8, 2010; Sydney, Australia p. 944-947. [doi: 10.1111/j.1442-9071.2010.02415.x]
50. Kononowicz AA, Krawczyk P, Cebula G, Dembkowska M, Drab E, Fraczek B, et al. Effects of introducing a voluntary virtual patient module to a basic life support with an automated external defibrillator course: a randomised trial. BMC Med Educ 2012 Jun 18;12:41 [FREE Full text] [doi: 10.1186/1472-6920-12-41] [Medline: 22709278]
51. Kumta SM, Tsang PL, Hung LK, Cheng JCY. Fostering critical thinking skills through a web-based tutorial programme for final year medical students--A randomized controlled study. J Educ Multimed Hypermedia 2003;12(3):267-273.
52. Succar T, Zebington G, Billson F, Byth K, Barrie S, McCluskey P, et al. The impact of the Virtual Ophthalmology Clinic on medical students' learning: a randomised controlled trial. Eye (Lond) 2013 Oct;27(10):1151-1157. [doi: 10.1038/eye.2013.143] [Medline: 23867718]
53. 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]
54. Bonnetain E, Boucheix JM, Hamet M, Freysz M. Benefits of computer screen-based simulation in learning cardiac arrest procedures. Med Educ 2010 Jul;44(7):716-722. [doi: 10.1111/j.1365-2923.2010.03708.x] [Medline: 20636591]
55. Botezatu M, Hult H, Tessma MK, Fors UGH. Virtual patient simulation for learning and assessment: superior results in comparison with regular course exams. Med Teach 2010;32(10):845-850. [doi: 10.3109/01421591003695287] [Medline: 20854161]
56. Botezatu M, Hult H, Tessma MK, Fors U. Virtual patient simulation: knowledge gain or knowledge loss? Med Teach 2010;32(7):562-568. [doi: 10.3109/01421590903514630] [Medline: 20653378]
57. Fleetwood J, Vaught W, Feldman D, Gracely E, Kassutto Z, Novack D. MedEthEx Online: a computer-based learning program in medical ethics and communication skills. Teach Learn Med 2000;12(2):96-104. [doi: 10.1207/S15328015TLM1202_7] [Medline: 11228685]
58. Haerling KA. Cost-utility analysis of virtual and mannequin-based simulation. Simul Healthc 2018 Feb;13(1):33-40. [doi: 10.1097/SIH.0000000000000280] [Medline: 29373382]
59. Jeimy S, Wang JY, Richardson L. Evaluation of virtual patient cases for teaching diagnostic and management skills in internal medicine: a mixed methods study. BMC Res Notes 2018 Jun 5;11(1):357 [FREE Full text] [doi: 10.1186/s13104-018-3463-x] [Medline: 29871699]

60. Kandasamy T, Fung K. Interactive internet-based cases for undergraduate otolaryngology education. Otolaryngol Head Neck Surg 2009 Mar;140(3):398-402. [doi: 10.1016/j.otohns.2008.11.033] [Medline: 19248951]
61. Kinney P, Keskula DR, Perry JF. The effect of a computer assisted instructional program on physical therapy students. J Allied Health 1997;26(2):57-61. [Medline: 9268782]
62. Leong SL, Baldwin CD, Adelman AM. Integrating web-based computer cases into a required clerkship: development and evaluation. Acad Med 2003 Mar;78(3):295-301. [Medline: 12634211]
63. 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]
64. Liaw SY, Chan SW, Chen F, Hooi SC, Siau C. Comparison of virtual patient simulation with mannequin-based simulation for improving clinical performances in assessing and managing clinical deterioration: randomized controlled trial. J Med Internet Res 2014;16(9):e214 [FREE Full text] [doi: 10.2196/jmir.3322] [Medline: 25230684]
65. Maleck M, Fischer MR, Kammer B, Zeiler C, Mangel E, Schenk F, et al. Do computers teach better? A media comparison study for case-based teaching in radiology. Radiographics 2001;21(4):1025-1032. [doi: 10.1148/radiographics.21.4.g01jl091025] [Medline: 11452078]
66. Miedzybrodzka Z, Hamilton NM, Gregory H, Milner B, Frade I, Sinclair T, et al. Teaching undergraduates about familial breast cancer: comparison of a computer assisted learning (CAL) package with a traditional tutorial approach. Eur J Hum Genet 2001 Dec;9(12):953-956 [FREE Full text] [doi: 10.1038/sj.ejhg.5200751] [Medline: 11840198]
67. Qayumi AK, Kurihara Y, Imai M, Pachev G, Seo H, Hoshino Y, et al. Comparison of computer-assisted instruction (CAI) versus traditional textbook methods for training in abdominal examination (Japanese experience). Med Educ 2004 Oct;38(10):1080-1088. [doi: 10.1111/j.1365-2929.2004.01957.x] [Medline: 15461653]
68. Schwid HA, Rooke GA, Ross BK, Sivarajan M. Use of a computerized advanced cardiac life support simulator improves retention of advanced cardiac life support guidelines better than a textbook review. Crit Care Med 1999 Apr;27(4):821-824. [Medline: 10321676]
69. Schwid HA, Rooke GA, Michalowski P, Ross BK. Screen-based anesthesia simulation with debriefing improves performance in a mannequin-based anesthesia simulator. Teach Learn Med 2001;13(2):92-96. [doi: 10.1207/S15328015TLM1302_4] [Medline: 11302037]
70. Secomb J, McKenna L, Smith C. The effectiveness of simulation activities on the cognitive abilities of undergraduate third-year nursing students: a randomised control trial. J Clin Nurs 2012 Dec;21(23-24):3475-3484. [doi: 10.1111/j.1365-2702.2012.04257.x] [Medline: 23145517]
71. Sobocan M, Turk N, Dinevski D, Hojs R, Balon BP. 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]
72. Subramanian A, Timberlake M, Mittakanti H, Lara M, Brandt ML. Novel educational approach for medical students: improved retention rates using interactive medical software compared with traditional lecture-based format. J Surg Educ 2012;69(2):253-256. [doi: 10.1016/j.jsurg.2011.12.007] [Medline: 22365876]
73. Tao H. Computer-based simulative training system—a new approach to teaching pre-hospital trauma care. J Med Coll PLA 2011 Dec;26(6):335-344. [doi: 10.1016/S1000-1948(12)60029-X]
74. Triola M, Feldman H, Kalet AL, Zabar S, Kachur EK, Gillespie C, et al. A randomized trial of teaching clinical skills using virtual and live standardized patients. J Gen Intern Med 2006 May;21(5):424-429 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00421.x] [Medline: 16704382]
75. Vash JH, Yunesian M, Shariati M, Keshvari A, Harirchi I. Virtual patients in undergraduate surgery education: a randomized controlled study. ANZ J Surg 2007;77(1-2):54-59. [doi: 10.1111/j.1445-2197.2006.03978.x] [Medline: 17295822]
76. Wang CL, Chinnugounder S, Hippe DS, Zaidi S, O'Malley RB, Bhargava P, et al. Comparative effectiveness of hands-on versus computer simulation-based training for contrast media reactions and teamwork skills. J Am Coll Radiol 2017 Jan;14(1):103-10.e3. [doi: 10.1016/j.jacr.2016.07.013] [Medline: 27815053]
77. Williams C, Aubin S, Harkin P, Cottrell D. A randomized, controlled, single-blind trial of teaching provided by a computer-based multimedia package versus lecture. Med Educ 2001 Sep;35(9):847-854. [Medline: 11555222]
78. Bryant R, Miller CL, Henderson D. Virtual clinical simulations in an online advanced health appraisal course. Clin Simul Nurs 2015 Oct;11(10):437-444. [doi: 10.1016/j.ecns.2015.08.002]
79. Deladisma AM, Gupta M, Kotranza A, Bittner JG, Imam T, Swinson D, et al. A pilot study to integrate an immersive virtual patient with a breast complaint and breast examination simulator into a surgery clerkship. Am J Surg 2009 Jan;197(1):102-106. [doi: 10.1016/j.amjsurg.2008.08.012] [Medline: 19101251]
80. Gu Y, Zou Z, Chen X. The Effects of vSIM for Nursing™ as a teaching strategy on fundamentals of nursing education in undergraduates. Clin Simul Nurs 2017 Apr;13(4):194-197. [doi: 10.1016/j.ecns.2017.01.005]
81. Kaltman S, Talisman N, Pennestri S, Syverson E, Arthur P, Vovides Y. Using technology to enhance teaching of patient-centered interviewing for early medical students. Simul Healthc 2018 Jun;13(3):188-194. [doi: 10.1097/SIH.0000000000000304] [Medline: 29771814]

82. Lehmann R, Thiessen C, Frick B, Bosse HM, Nikendei C, Hoffmann GF, et al. Improving pediatric basic life support performance through blended learning with web-based virtual patients: randomized controlled trial. J Med Internet Res 2015 Jul 2;17(7):e162 [FREE Full text] [doi: 10.2196/jmir.4141] [Medline: 26139388]
83. Schittek Janda M, Mattheos N, Nattestad A, Wagner A, Nebel D, Färbom C, et al. Simulation of patient encounters using a virtual patient in periodontology instruction of dental students: design, usability, and learning effect in history-taking skills. Eur J Dent Educ 2004 Aug;8(3):111-119. [doi: 10.1111/j.1600-0579.2004.00339.x] [Medline: 15233775]
84. Smith BD, Silk K. Cultural competence clinic: an online, interactive, simulation for working effectively with Arab American Muslim patients. Acad Psychiatry 2011;35(5):312-316. [doi: 10.1176/appi.ap.35.5.312] [Medline: 22007089]
85. Wahlgren CF, Edelbring S, Fors U, Hindbeck H, Ståhle M. Evaluation of an interactive case simulation system in dermatology and venereology for medical students. BMC Med Educ 2006;6:40 [FREE Full text] [doi: 10.1186/1472-6920-6-40] [Medline: 16907972]
86. Weverling GJ, Stam J, ten Cate TJ, van Crevel H. [Computer-assisted education in problem-solving in neurology; a randomized educational study]. Ned Tijdschr Geneeskd 1996 Feb 24;140(8):440-443. [Medline: 8720819]
87. Courteille O, Fahlstedt M, Ho J, Hedman L, Fors U, von Holst H, et al. Learning through a virtual patient vs recorded lecture: a comparison of knowledge retention in a trauma case. Int J Med Educ 2018 Mar 28;9:86-92 [FREE Full text] [doi: 10.5116/ijme.5aa3.ccf2] [Medline: 29599421]
88. Dankbaar ME, Alsma J, Jansen EE, van Merrienboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. Adv Health Sci Educ Theory Pract 2016 Aug;21(3):505-521. [doi: 10.1007/s10459-015-9641-x] [Medline: 26433730]
89. Foster A, Chaudhary N, Murphy J, Lok B, Waller J, Buckley PF. The use of simulation to teach suicide risk assessment to health profession trainees—rationale, methodology, and a proof of concept demonstration with a virtual patient. Acad Psychiatry 2015 Dec;39(6):620-629. [doi: 10.1007/s40596-014-0185-9] [Medline: 25026950]
90. Trudeau KJ, Hildebrand C, Garg P, Chiauzzi E, Zacharoff KL. A randomized controlled trial of the effects of online pain management education on primary care providers. Pain Med 2017 Dec 1;18(4):680-692. [doi: 10.1093/pm/pnw271] [Medline: 28034967]
91. Bearman M, Cesnik B, Liddell M. Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. Med Educ 2001 Sep;35(9):824-832. [Medline: 11555219]
92. Berger J, Bawab N, de Mooij J, Widmer DS, Szilas N, de Vriese C, et al. An open randomized controlled study comparing an online text-based scenario and a serious game by Belgian and Swiss pharmacy students. Curr Pharm Teach Learn 2018 Dec;10(3):267-276. [doi: 10.1016/j.cptl.2017.11.002] [Medline: 29764629]
93. Braun LT, Zottmann JM, Adolf C, Lottspeich C, Then C, Wirth S, et al. Representation scaffolds improve diagnostic efficiency in medical students. Med Educ 2017 Nov;51(11):1118-1126. [doi: 10.1111/medu.13355] [Medline: 28585351]
94. Davids MR, Chikte UM, Halperin ML. Effect of improving the usability of an e-learning resource: a randomized trial. Adv Physiol Educ 2014 Jun;38(2):155-160 [FREE Full text] [doi: 10.1152/advan.00119.2013] [Medline: 24913451]
95. Foster A, Chaudhary N, Kim T, Waller JL, Wong J, Borish M, et al. Using virtual patients to teach empathy: a randomized controlled study to enhance medical students' empathic communication. Simul Healthc 2016 Jun;11(3):181-189. [doi: 10.1097/SIH.0000000000000142] [Medline: 26841278]
96. Friedman CP, France CL, Drossman DD. A randomized comparison of alternative formats for clinical simulations. Med Decis Making 1991;11(4):265-272. [doi: 10.1177/0272989X9101100404] [Medline: 1766329]
97. Harris JM, Sun H. A randomized trial of two e-learning strategies for teaching substance abuse management skills to physicians. Acad Med 2013 Sep;88(9):1357-1362 [FREE Full text] [doi: 10.1097/ACM.0b013e31829e7ec6] [Medline: 23887001]
98. 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]
99. Maier EM, Hege I, Muntau AC, Huber J, Fischer MR. What are effects of a spaced activation of virtual patients in a pediatric course? BMC Med Educ 2013 Mar 28;13:45 [FREE Full text] [doi: 10.1186/1472-6920-13-45] [Medline: 23537162]
100. Tolsgaard MG, Jepsen RMHG, Rasmussen MB, Kayser L, Fors U, Laursen LC, et al. The effect of constructing versus solving virtual patient cases on transfer of learning: a randomized trial. Perspect Med Educ 2016 Feb;5(1):33-38 [FREE Full text] [doi: 10.1007/s40037-015-0242-4] [Medline: 26754313]
101. Kraiger K, Ford JK, Salas E. Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. J Appl Psychol 1993;78(2):311-328. [doi: 10.1037/0021-9010.78.2.311] [Medline: 31070385]
102. Stevens A, Hernandez J, Johnsen K, Dickerson R, Raij A, Harrison C, et al. The use of virtual patients to teach medical students history taking and communication skills. Am J Surg 2006 Jun;191(6):806-811. [doi: 10.1016/j.amjsurg.2006.03.002] [Medline: 16720154]
103. Friedman CP. The research we should be doing. Acad Med 1994 Jun;69(6):455-457. [Medline: 8003158]
104. 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]
105. Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theory Pract 2002 Nov;41(4):212-218. [doi: 10.1207/s15430421tip4104_2]

106. Chen F, Lui AM, Martinelli SM. A systematic review of the effectiveness of flipped classrooms in medical education. Med Educ 2017 Jun;51(6):585-597. [doi: 10.1111/medu.13272] [Medline: 28488303]
107. Kirschner PA, Sweller J, Clark RE. Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educ Psychol 2006 Jun;41(2):75-86. [doi: 10.1207/s15326985ep4102_1]
108. Edelbring S, Wahlström R. Dynamics of study strategies and teacher regulation in virtual patient learning activities: a cross sectional survey. BMC Med Educ 2016 Apr 23;16:122 [FREE Full text] [doi: 10.1186/s12909-016-0644-y] [Medline: 27108089]

## Abbreviations

**3D:** 3-dimensional
**cRCT:** cluster randomized controlled trials
**GRADE:** Grading of Recommendations Assessment, Development and Evaluation Working Group
**MCQ:** multiple-choice question
**OSCE:** objective structured clinical examination
**RCT:** randomized controlled trial
**SMD:** standardized mean difference

*Edited by G Eysenbach, A Marusic; submitted 10.05.19; peer-reviewed by SY Liaw, M Sobocan; comments to author 31.05.19; revised version received 06.06.19; accepted 08.06.19; published 02.07.19.*

*Please cite as:*
*Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, Tudor Car L, Carlstedt-Duke J, Car J, Zary N*
*Virtual Patient Simulations in Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration*
*J Med Internet Res 2019;21(7):e14676*
*URL: https://www.jmir.org/2019/7/e14676/*
*doi:10.2196/14676*
*PMID:31267981*

Corrigenda and Addenda

# Multimedia Appendix Correction: Technological Innovations in Disease Management: Text Mining US Patent Data From 1995 to 2017

Ming Huang[1], PhD; Maryam Zolnoori[1], PhD; Joyce E Balls-Berry[1], MPE, PhD; Tabetha A Brockman[2,3], MA; Christi A Patten[2,3], PhD; Lixia Yao[1], PhD

[1]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States

[2]Center for Clinical and Translational Science, Commuity Engagement Program, Mayo Clinic, Rochester, MN, United States

[3]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, United States

**Corresponding Author:**
Lixia Yao, PhD
Department of Health Sciences Research
Mayo Clinic
200 First Street SW
Rochester, MN, 55905
United States
Phone: 1 507 293 7953
Fax: 1 507 284 1516
Email: Yao.Lixia@mayo.edu

**Related Article:**

Correction of: http://www.jmir.org/2019/4/e13316/

***(J Med Internet Res 2019;21(7):e14678)*** doi:10.2196/14678

The authors of "Technological Innovations in Disease Management: Text Mining US Patent Data From 1995 to 2017" (J Med Internet Res 2019;21(4):e13316) have provided a new version of Multimedia Appendix 3.

The originally published version of Multimedia Appendix 3 contained raw treatment costs of diseases as disease burden in Table S6 that were calculated with claims data from OptumLabs. The sharing of raw data violates the data policy of OptumLabs, so those raw numbers have been removed.

The correction will appear in the online version of the paper on the JMIR website on July 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.

## Multimedia Appendix 3

Patent coverage (Table S5), ROIs (Table S6), PHIs (Table S7), and topics (Tables S8 and S9) for over 600 diseases and medical conditions.

[XLSX File (Microsoft Excel File), 25MB - jmir_v21i7e14678_app3.xlsx ]

*Submitted 10.05.19; this is a non–peer-reviewed article;accepted 13.06.19; published 05.07.19.*

*Please cite as:*
*Huang M, Zolnoori M, Balls-Berry JE, Brockman TA, Patten CA, Yao L*
*Multimedia Appendix Correction: Technological Innovations in Disease Management: Text Mining US Patent Data From 1995 to 2017*
*J Med Internet Res 2019;21(7):e14678*
*URL: http://www.jmir.org/2019/7/e14678/*
*doi:10.2196/14678*
*PMID:31278733*

Corrigenda and Addenda

# Metadata and Table Caption Correction: What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews

Y Alicia Hong[1,2], PhD; Chen Liang[3], PhD; Tiffany A Radcliff[2], PhD; Lisa T Wigfall[4], PhD; Richard L Street[5], PhD

[1]Department of Health Administration and Policy, George Mason University, Fairfax, VA, United States
[2]School of Public Health, Texas A&M University, College Station, TX, United States
[3]Arnold School of Public Health, University of South Carolina, Columbia, SC, United States
[4]Department of Health Kinesiology, Texas A&M University, College Station, TX, United States
[5]Department of Communication, Texas A&M University, College Station, TX, United States

**Corresponding Author:**
Y Alicia Hong, PhD
Department of Health Administration and Policy
George Mason University
4400 University Drive, MS 1J3
Fairfax, VA, 22030
United States
Phone: 1 703 993 1929
Email: yhong22@gmu.edu

**Related Article:**

Correction of: http://www.jmir.org/2019/4/e12521/

***(J Med Internet Res 2019;21(7):e14823)*** doi:10.2196/14823

The authors of "What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews" (J Med Internet Res 2019;21(4):e12521) made an error in the caption of Table 2. It previously read "Summaries of published studies on patient online reviews (63 studies consisting of 69 articles)" but has now been changed to "Studies that compare patient online reviews with traditional healthcare quality indicators".

The lead author, Y Alicia Hong, now has an additional affiliation (Department of Health Administration and Policy, George Mason University, Fairfax, VA, United States) in addition to her previous affiliation (School of Public Health, Texas A&M University, College Station, TX, United States). This has bumped the numbering of all other affiliations by one, although affiliations remain the same for all other authors.

Y Alicia Hong is also the corresponding author and wishes to use her new affiliation's information for correspondence. The previous contact information was as follows:

*Y Alicia Hong, PhD*
*School of Public Health*
*Texas A&M University*
*212 Adriance Lab Road*
*College Station, TX, 77843-1266*
*United States*
*Phone: 1 9794369343*
*Email: yhong@sph.tamhsc.edu*

The new contact information is as follows:

*Y Alicia Hong, PhD*
*Department of Health Administration and Policy*
*George Mason University 4400 University Drive, MS 1J3*
*Fairfax, VA, 22030*
*United States*
*Phone: 1 703 993 1929*
*Email: yhong22@gmu.edu*

The correction will appear in the online version of the paper on the JMIR website on July 18, 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.

*Submitted 26.05.19; this is a non–peer-reviewed article;accepted 13.06.19; published 18.07.19.*

*Please cite as:*
*Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL*
*Metadata and Table Caption Correction: What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews*
*J Med Internet Res 2019;21(7):e14823*
*URL: http://www.jmir.org/2019/7/e14823/*
*doi:10.2196/14823*
*PMID:31322126*

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org