Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 20 (2018), Issue 7 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Tutorial

Review

Patient-Centered eHealth Interventions for Children, Adolescents, and Adults With Sickle Cell Disease: Systematic Review

Sherif M Badawy[1,2,3*], MS, MD, MBBCh; Robert M Cronin[4,5*], MS, MD; Jane Hankins[6], MS, MD; Lori Crosby[7,8], PsyD; Michael DeBaun[9], MPH, MD; Alexis A Thompson[1,2], MPH, MD; Nirmish Shah[10], MD

[1]Division of Hematology, Oncology and Stem Cell Transplant, Ann & Robert H Lurie Children's Hospital of Chicago, Chicago, IL, United States

[2]Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Department of Pediatrics, Division of Hematology and Oncology, Faculty of Medicine, Zagazig University, Zagazig, Egypt

[4]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[5]Department of Internal Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[6]Department of Hematology, St Jude Children's Research Hospital, Memphis, TN, United States

[7]Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[8]Department of Psychology, University of Cincinnati, Cincinnati, OH, United States

[9]Division of Hematology and Oncology, Department of Pediatrics, Vanderbilt-Meharry Center for Excellence in Sickle Cell Disease, Vanderbilt University Medical Center, Nashville, TN, United States

[10]Division of Hematology, Duke University School of Medicine, Durham, NC, United States

[*]these authors contributed equally

Corresponding Author:
Sherif M Badawy, MS, MD, MBBCh
Division of Hematology, Oncology and Stem Cell Transplant
Ann & Robert H Lurie Children's Hospital of Chicago
225 E Chicago Ave, Box #30
Chicago, IL, 60611
United States
Phone: 1 3122274836
Fax: 1 3122279373
Email: sbadawy@luriechildrens.org

Abstract

Background: Sickle cell disease is an inherited blood disorder that affects over 100,000 Americans. Sickle cell disease–related complications lead to significant morbidity and early death. Evidence supporting the feasibility, acceptability, and efficacy of self-management electronic health (eHealth) interventions in chronic diseases is growing; however, the evidence is unclear in sickle cell disease.

Objective: We systematically evaluated the most recent evidence in the literature to (1) review the different types of technological tools used for self-management of sickle cell disease, (2) discover and describe what self-management activities these tools were used for, and (3) assess the efficacy of these technologies in self-management.

Methods: We reviewed literature published between 1995 and 2016 with no language limits. We searched MEDLINE, EMBASE, CINAHL, PsycINFO, and other sources. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. Two independent reviewers screened titles and abstracts, assessed full-text articles, and extracted data from articles that met inclusion criteria. Eligible studies were original research articles that included texting, mobile phone–based apps, or other eHealth interventions designed to improve self-management in pediatric and adult patients with sickle cell disease.

Results: Of 1680 citations, 16 articles met all predefined criteria with a total of 747 study participants. Interventions were text messaging (4/16, 25%), native mobile apps (3/16, 19%), Web-based apps (5/16, 31%), mobile directly observed therapy (2/16, 13%), internet-delivered cognitive behavioral therapy (2/16, 13%), electronic pill bottle (1/16, 6%), or interactive gamification (2/16, 13%). Interventions targeted monitoring or improvement of medication adherence (5/16, 31%); self-management, pain reporting, and symptom reporting (7/16, 44%); stress, coping, sleep, and daily activities reporting (4/16, 25%); cognitive training for memory (1/16, 6%); sickle cell disease and reproductive health knowledge (5/16, 31%); cognitive behavioral therapy (2/16, 13%); and guided relaxation interventions (1/16, 6%). Most studies (11/16, 69%) included older children or adolescents (mean

or median age 10-17 years; 11/16, 69%) and 5 included young adults (≥18 years old) (5/16, 31%). Sample size ranged from 11 to 236, with a median of 21 per study: <20 in 6 (38%), ≥20 to <50 in 6 (38%), and >50 participants in 4 studies (25%). Most reported improvement in self-management–related outcomes (15/16, 94%), as well as high satisfaction and acceptability of different study interventions (10/16, 63%).

Conclusions: Our systematic review identified eHealth interventions measuring a variety of outcomes, which showed improvement in multiple components of self-management of sickle cell disease. Despite the promising feasibility and acceptability of eHealth interventions in improving self-management of sickle cell disease, the evidence overall is modest. Future eHealth intervention studies are needed to evaluate their efficacy, effectiveness, and cost effectiveness in promoting self-management in patients with sickle cell disease using rigorous methods and theoretical frameworks with clearly defined clinical outcomes.

(J Med Internet Res 2018;20(7):e10940) doi:10.2196/10940

KEYWORDS

sickle cell; self-management; eHealth; mHealth; interventions; internet; anemia, sickle cell; telemedicine

Introduction

Background

Sickle cell disease (SCD) is an inherited blood disorder that affects more than 5 million individuals worldwide, and about 250,000 babies with SCD are born every year, mainly in Africa [1]. SCD affects over 100,000 Americans, mainly African Americans, many of whom are of lower socioeconomic status [2-5]. Advancements in treatment over the past few decades have changed the course of SCD from a condition of childhood to a chronic disease of adulthood [6]. Individuals with SCD are subject to acute and chronic complications, including vaso-occlusive pain crisis, acute chest syndrome, stroke, cognitive dysfunction, and end-organ damage to the liver, spleen, and kidneys, substantially reducing health-related quality of life and leading to early death [7,8]. Management of these SCD complications has a significant impact on health care utilization in the United States, with over 230,000 emergency room visits per year with an annual health expenditures of US $1.5 billion [9,10].

As part of the chronic care model [11], creating the informed, activated patient, along with the proactive care team, can lead to better health outcomes. One essential component of the informed, activated patient is the concept of self-management. Self-management has been referred to as the individual's ability to manage the symptoms, treatment, physical and psychological consequences, and lifestyle changes inherent in living with a chronic condition [12]. SCD patients with more self-management skills can better manage their illness and potentially improve their health outcomes. Self-management skills are key for SCD patients as they encounter challenges related to managing their illness, such as pain management, adequate hydration, balanced nutrition, clinic attendance, and adherence to medication regimens, especially after they transition from pediatric to adult care settings. In particular, medication adherence is an important component of self-management. SCD patients with more reported adherence barriers and negative perceptions of hydroxyurea or SCD reported lower adherence rates and worse health-related quality of life scores [13-15]. In addition, many SCD patients were interested in having mobile apps with up-to-date clinical care guidelines that can improve understanding of the importance of self-management [16] and apps with features to improve their disease self-management [17]. Different techniques have been used to foster greater self-management and involve nontechnological solutions (eg, face-to-face or paper-based interventions). Over the last two decades, however, technological solutions, especially using the internet and mobile phones, to improve self-management have gained momentum with the wide access to mobile devices. These solutions, particularly electronic health (eHealth) interventions, potentially provide the benefit of greater acceptance and dissemination. eHealth has been defined as "an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the internet and related technologies" [18].

Access to personal technology is ubiquitous, and technological solutions are becoming a part of the way health care is delivered. People are more frequently using technology for their health [19,20], and there are government mandates, including meaningful use in the United States [21], that require health care providers to use technology in the care of their patients. Moreover, enhanced patient activation, as well as engagement in medical care and shared decision making, has been associated with improved health outcomes [22-24]. eHealth interventions have been shown to improve patient activation and engagement [25-29], making them a possible solution to improve outcomes. In addition, individuals with SCD and their families want to use technologies for their health [17,30,31]. While some eHealth technological interventions are being created and tested in SCD, these interventions have not been sufficiently evaluated in the few existing studies. Furthermore, a discussion about what interventions exist, how efficacious they are, and how they are being used to improve disease self-management is missing in this population.

In other chronic diseases, such as diabetes and asthma, a growing body of evidence has described improvements in self-care through the use of technological interventions [32-34]. Additionally, recent systematic reviews showed promising data to support the overall feasibility, acceptability, and efficacy of mobile interventions in improving health outcomes [35-40], although cost effectiveness remains unclear [41]. However, to the best of our knowledge, there has not been a systematic review of technological interventions used to improve self-management in the care of SCD. Further, evidence is growing to support the benefits of using mobile interventions

to improve self-management in patients with chronic health conditions living in low- and middle-income countries [42-49]. Given the broad access to personal technology, as well as the high prevalence of SCD in many African countries, developing and implementing evidence-based mobile interventions could provide an opportunity to improve self-management skills and health outcomes in this population.

Objectives

To understand how eHealth technology has been used to increase self-management of SCD and to guide future research, we performed a systematic review of the literature with the following objectives: (1) to review the different types of technological tools used for self-management of SCD, (2) to discover and describe what self-management activities these tools were used for, and (3) to assess the efficacy of these technologies in self-management. We conclude with gaps that will need to be addressed in future research.

Methods

Literature Search

This systematic review covered literature published between 1995 and 2016 with no language limits. The search strategy looked for all articles on texting, phones, mobile phone apps, portable software, and other eHealth interventions combined with sickle cell search terms. We intentionally used the Boolean operator OR instead of AND to capture the most comprehensive set of articles possible to which we applied our eligibility criteria. In brief, a medical librarian conducted the literature search in the following sources from inception to August 30, 2016: MEDLINE (through PubMed), EMBASE, Web of Science, Cochrane Central Register of Controlled Trials, CINAHL, PsycINFO, Engineering Village, and ClinicalTrials.gov databases. After the initial search, the results of the search were limited to articles published from 1995 to the date of the search on November 22, 2016. Our search strategy began with the MEDLINE search and was translated to the appropriate syntax for each of the other databases. In addition, we hand searched related themes. Multimedia Appendix 1 shows the detailed search strategies. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines in the reporting of evidence across the studies reviewed herein (Multimedia Appendix 2) [50]. Two independent reviewers (SMB and RMC) assessed abstracts and articles against the eligibility criteria and critically appraised the methodological quality using established criteria from the Centre for Evidence-based Medicine [51]. Disagreements were resolved by discussion or consultation with a colleague, if needed.

Eligibility Criteria

Eligible studies were original research articles reporting randomized controlled trials (RCTs), quasi-experimental studies, or pilot pre-post studies of texting, mobile phone–based apps, or other eHealth interventions designed to improve self-management in pediatric and adult patients with SCD. To be included in this review, the studies had to report at least one primary or secondary outcome related to self-management behavior, such as medication adherence, pain management, or education. We excluded studies focused on physicians or providers, or other aspects of SCD care other than self-management.

Data Synthesis

We used a standardized form for data extraction. Data items in the extraction form were the following: first author's name; publication year; country; aim of the study; participants' age and sex; study design and setting; sample size; selection criteria; duration of intervention and follow-up; retention rate; components of the study intervention (texting, mobile phone apps, or other eHealth interventions) and comparator (if applicable); self-management measures and outcomes; other related outcome; and theoretical framework.

Results

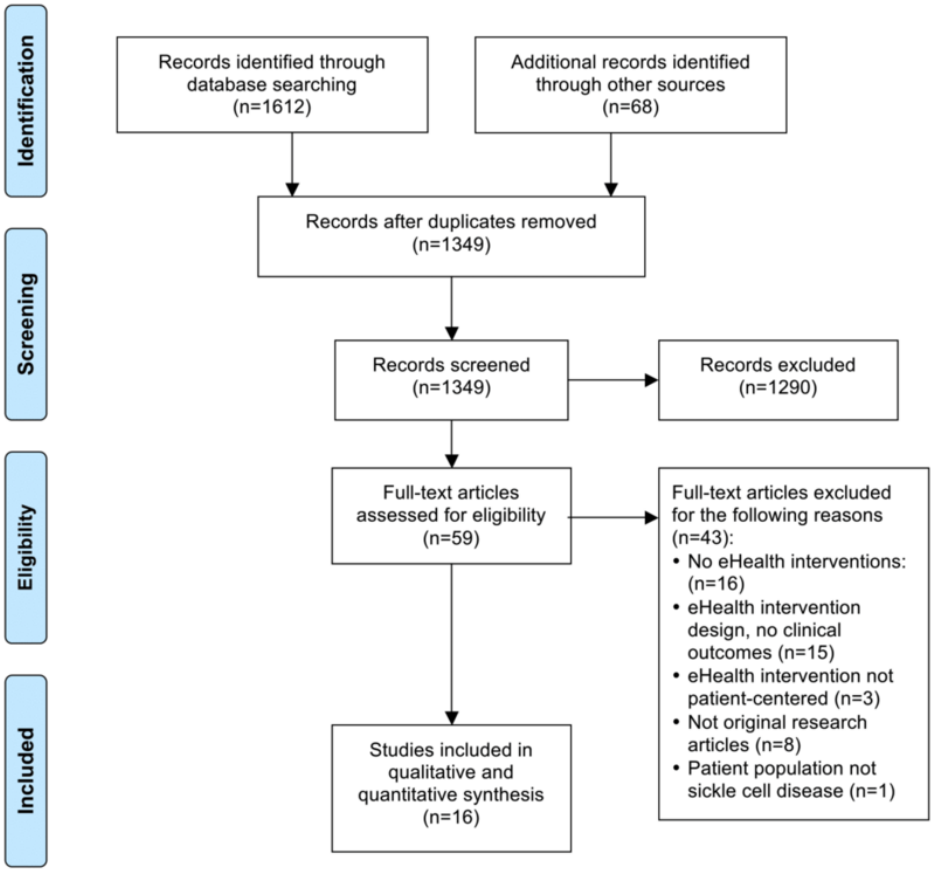
Literature Search

We retrieved a total of 1680 citations: 1612 identified through searching different databases and 68 through other resources. After we removed duplicates, 1349 original articles remained (Figure 1). Two authors (SMB and RMC) independently screened the article titles and abstracts of the 1349 records against the inclusion criteria, and a total of 59 met all predefined inclusion criteria. The same 2 authors (SMB and RMC) then independently reviewed the full text of these articles in detail against the exclusion criteria and excluded 43 articles. Finally, 16 articles met all predefined criteria to be included in this review with a total of 747 study participants [52-67]. We did not identify any non-English articles that met our inclusion criteria. Figure 1 shows the study PRISMA flowchart and documents the reasons for exclusion of full-text articles.

Description of Included Studies

Multimedia Appendix 3 [52-67] summarizes the studies' characteristics. The aims of the interventions were (1) monitoring or improvement, or both, of medication adherence, including hydroxyurea [53,54,58,63], iron chelation [61], or asthma medications [63]; (2) self-management [52,59,60,62,64,66]; (3) pain and symptom reporting [52,55,59,60,62,64]; (4) stress, coping, sleep, and daily activities reporting [55,59,62,64]; (5) cognitive training for memory [57]; (6) disease education to improve SCD and [56,61,65-67] reproductive health knowledge [56,65]; and (7) cognitive behavioral therapy [62,64] and guided relaxation interventions [55]. All studies were performed in the United States [52-67]. Enrollment was mainly from clinics [52-58,60-66], as well as inpatients [67], the public [65], online networks [65], home [56,65], or community-based organizations [56,59,65]. All studies were conducted in the outpatient setting [52-66], except for 1 study conducted in the inpatient setting [67]. Most studies included older children or adolescents (mean or median age 10-17 years) [52-54,57-59,61,62,64,66,67], 5 studies included young adults (≥18 years old) [55,56,60,63,65], and 2 studies allowed parents to participate [54,58]. None of the included studies involved potential users, patients, or parents in the development of the intervention before it was tested.

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



Sample size ranged from 11 to 236, with a median of 21 per study; 6 enrolled fewer than 20 participants [53,57,58,60-62], 6 had 20 to 50 participants [52,55,56,63,64,66], and 4 enrolled more than 50 participants [54,59,65,67].

Description of Study Methodologies

Study design varied in the included studies: 7 were pre-post pilot or feasibility trials [53,56-58,60,61,66], 5 were RCTs [55,62-65], 2 were single-arm observational studies [52,59], 1 was a quasi-experimental study [67], and 1 was a retrospective study [54]. Of the RCTs, 3 were nonblinded [62-64], 1 was single-blinded (participants) [65], and 1 was double-blinded [55]. Details of allocation concealment and study blinding were not adequately reported. None of the RCTs used intention-to-treat analysis. Retention rates differed across studies: less than 80% in 4 studies [52,60,64,67], from 80% to less than 100% in 6 studies [55,58,61-63,65], 100% in 1 study [53], and not reported in 5 studies [54,56,59,66,67]. The duration of the intervention ranged from 3 days to 12 months as follows: 3 months or less [55-57,60,62,63,67], more than 3 to 6 months or less [53,58,59,61,64], or more than 6 up to 12 months [52,54]. A total of 3 studies implemented a reward system to enhance participant engagement during the study intervention [52,53,67], and 1 study assessed the sustainability of the intervention effects with 3-month follow-up after the completion of the active intervention [61]. Additionally, 6 studies were informed by a clear theoretical framework for their intervention effects, as follows: health belief model [63]; theory-based game design [66]; gate control theory [55]; transactional stress model [67]; coping theoretical model [67]; theory of reasoned action [56,65]; and Kolb experiential learning theory [56,65].

Description of eHealth Interventions

Interventions included text messaging [52-54,59,63], native mobile apps [57,60,61], Web-based apps [52,55,56,59,65], mobile directly observed therapy [53,61], internet-delivered cognitive behavioral therapy [62,64], electronic pill bottle [58], or interactive gamification [66,67]. A total of 4 studies indicated regular or mobile phone ownership or access as a requirement for study participation [52-54,63], while loaner phones [59,62,64], loaner tablets [55-57,65], or both [60,61] were available in other studies. Multimedia Appendix 4 [52-67] summarizes the various intervention components for all included studies.

Intervention Effects on Outcomes

Study outcomes varied across studies, including medication adherence [53,54,58,61,63], disease knowledge [56,61,65-67], reproductive health knowledge [56,65], pain or symptom reporting completion rates [52,55,59,60,62,64], health care utilization [54,59], total opioid use [55], self-management skills [52,59,62,64,66], and coping and social support [67]. Almost all studies (15/16, 94%) reported improvement in self-management outcomes [52-57,59-67]. Most (10/16, 63%) reported high satisfaction and acceptability of different study interventions [52,53,55,56,59-62,65,67], while 6 studies did not report on these outcomes. Table 1 summarizes the main outcomes for all included studies.

Table 1. Summary of the main outcomes for all included studies with eHealth interventions.

Study	Main outcome
Jonassaint, 2015 [60]	High correlation between paper and electronic (SMART[a] app) pain measurements; high association between pain severity and pain intensity using SMART app; daily entries using SMART app entries: 86% in week 1 and 58% in week 4; higher rates of daily entries with iPads and patients >35 years old; high usability and acceptability as a tool to monitor daily pain and other symptoms.
Hardy, 2016 [57]	Average number of completed sessions was 15.83 (SD 7.73); participants with higher completion rates were female and had lower pain scores; participants who completed scheduled intervention (Cogmed) sessions had improved verbal working memory, as well as visuospatial short-term and working memories.
Leonard, 2017 [61]	Participants tracked their medication usage about 80% at 30- and 90-day follow-up; high disease knowledge retention; adherence to iron chelation improved at 6-month follow-up as measured by serum ferritin levels and medication possession ratio; high satisfaction and acceptability as a tool to monitor medication adherence.
Creary, 2014 [53]	Adherence to hydroxyurea improved at 6-month follow-up as measured by fetal hemoglobin, mean corpuscular volume, and medication possession ratio; high satisfaction with electronic directly observed therapy (e-DOT) as a tool to monitor medication adherence; e-DOT needed 5 minutes or less to complete every day.
Estepp, 2014 [54]	Adherence to hydroxyurea improved as measured by laboratory markers (hemoglobin, fetal hemoglobin, mean corpuscular volume, absolute reticulocyte counts bilirubin levels); adherence to hydroxyurea improved as measured by medication possession ratio; no noticeable change in the number of hospitalizations or emergency room visits.
Pernell, 2017 [63]	Response rate to daily messages varied and was overall <50%; medication adherence self-report improved in the intervention group, but not in controls; asthma control test scores improved in the intervention group in adults, but not children.
Inoue, 2016 [58]	Hydroxyurea adherence rates were 85% as measured by either the electronic pill bottle GlowCap or medication possession ratio; laboratory markers of hydroxyurea adherence varied; a few technical challenges were also reported.
McClellan, 2009 [62]	Participants practiced I-CBT[b] coping skills with different frequencies; self-report practice rates were higher in older and male participants; high satisfaction as a tool for pain, sleep, coping, and daily activities reporting.
Schatz, 2015 [64]	Number of active psychological coping attempts increased with the intervention; reduction in pain scores when participants used I-CBT skills the day before for higher pain; no association between participants' skill use and functional activity.
Ezenwa, 2016 [55]	Intervention participants had significant reduction in current pain and stress levels; intervention participants had significant reduction in 2-week pain, but not stress intensity; no differences in total opioid use; high satisfaction with the tablet-based guided relaxation intervention to reduce pain.
Bakshi, 2017 [52]	Pain was reported most of the study days (76%); 50th and 90th percentiles of maximum daily pain directly correlated positively with mean maximum daily pain; proportion of pain-free days inversely correlated with mean maximum daily pain; highest pain diary completion rates were in first 30 days, which decreased over time; high satisfaction with momentary pain reporting and communication with medical team.
Jacob, 2013 [59]	Many children and adolescents reported mild to severe pain at home that did not require further evaluation by a health care professional; reported symptoms varied, including tiredness/fatigue, headache, yellowing of the eyes, and respiratory and musculoskeletal symptoms; higher pain scores were associated with shorter sleep duration and lower sleep quality; having previous history of SCD[c]-related events, symptoms, and negative thoughts was associated with reporting more frequent and higher-intensity pain; no differences in health care utilization (eg, emergency room visits or hospitalizations); high usability and acceptability as a tool to monitor daily pain and other symptoms.
Gallo, 2014 [56]	Intervention participants reported increased disease and reproductive knowledge scores; high acceptability of the CHOICES intervention; participants provided constructive feedback (eg, content, visualization, animation).
Wilkie, 2013 [65]	Intervention participants reported increased disease and reproductive knowledge scores; intervention participants were more likely to report a parenting plan to avoid SCD or SCD and sickle cell trait; there was an intervention effect on participants' parenting intention and planned behavior.
Hazzard, 2002 [67]	Participants' knowledge about SCD and asthma increased; participants reported more positive perceptions of peer support and less negative coping.
Yoon, 2007 [66]	Participants' SCD knowledge and confidence levels increased significantly.

[a]SMART: Sickle cell disease Mobile Application to Record symptoms via Technology.

[b]I-CBT: internet-delivered cognitive behavioral therapy.

[c]SCD: sickle cell disease.

Discussion

Principal Findings

eHealth is increasingly being used for self-management of a variety of chronic diseases, including asthma, diabetes, and hypertension [46], as well as SCD. Despite systematic reviews describing eHealth use in other chronic diseases, to our knowledge, this is the first systematic review of eHealth for self-management of SCD. Our review demonstrates a range of eHealth interventions measuring a variety of outcomes, which

showed improvement in multiple components of self-management of SCD. We also showed that few eHealth studies in SCD used rigorous methods, were grounded in theoretical frameworks, or measured clinical outcomes. This review describes the promise of eHealth to improve the care of individuals with SCD; however, many areas of future research can help demonstrate the usefulness of eHealth in this population.

Most studies were in children and adolescents with SCD. Many of these studies focused on adherence to medications such as hydroxyurea or iron chelation. Other systematic reviews looked at medication adherence using eHealth [68-73], with 1 of them looking specifically at the adolescent population [39]. Most of the studies in pediatric SCD had small sample sizes, and 1 was an RCT. These studies confirmed improvement in medication adherence in the participants receiving eHealth interventions. There is significant promise for improving medication adherence using eHealth in SCD, but larger, more methodologically rigorous studies demonstrating an improved effect are needed. Most of the other studies in our review focused on pain management and coping strategies in children. These studies also demonstrated good adherence to pain diaries and improved coping. One systematic review looked at the use of eHealth in pain [74], but the studies in this review were primarily in middle-aged participants. Another review described that studies in eHealth for pain in children are lacking [75]. The studies in this review exhibited the potential for eHealth interventions to improve self-management of pain in children with painful chronic diseases. Further, in our review, only 2 studies allowed caregivers to participate. Caregivers are an essential component of the care of the child, and more studies are needed to evaluate the use of eHealth in the parent-and-child dynamic to better understand optimal use of eHealth for both parts of this dynamic.

Intervention design did not vary according to patient characteristics, such as age, educational level, or other SCD-specific factors, which would be important for future intervention studies to consider as a strategy to optimize behavior change and long-term engagement. Additionally, 1 study was conducted in an inpatient setting, where management is more controlled by the health care system, whereas the goal of self-management interventions is to engage and empower patients in the outpatient setting with their day-to-day activities. In the outpatient setting, the health care system has less control, and the patient has more responsibility for disease management. More research is needed to evaluate the value of starting behavior change in the inpatient setting that could help to maintain intervention effects in the outpatient setting.

Relatively few studies evaluated eHealth in adults with SCD. This is in contrast to the number of systematic reviews of the use of eHealth in adults with other chronic diseases [33,46,76]. Most of the studies in adults with SCD focused on pain or knowledge about reproductive health in this age range, with only 1 study focused on medication adherence. However, the overall number of studies was fewer, and they were less concentrated on medication adherence, than the studies focused on eHealth use in children with SCD. More studies are needed in adults with SCD to demonstrate improvement in other components of care, including medication adherence for other medications, coping strategies, and clinic appointment adherence. Interestingly, half of the studies in this group were RCTs, which was more than those conducted in children and adolescents.

Studies in other chronic diseases measured outcomes unexplored in SCD. Multiple systematic reviews of eHealth in other chronic diseases saw improvements in clinical outcomes such as glycemic, blood pressure, and lipid level control [46,77]. None of the studies in our review evaluated the effect of eHealth interventions on outcomes for SCD, such as episodes of acute chest syndrome, strokes, or vaso-occlusive episodes of pain requiring emergency room visits or hospitalizations. Some of the pain studies in our review evaluated days and severity of pain, but these studies did not measure those pain episodes that resulted in health care utilization. Other reviews demonstrated improvement in clinic appointment adherence with eHealth interventions [76]. While there were preliminary studies in SCD describing an eHealth app to help with clinic appointments [78], there were no formal evaluation studies to demonstrate improvement in clinic attendance. Other studies looked at improving patient activation using eHealth in other diseases [79]; Risling and colleagues' review was primarily about patient portals that improve activation. Our review did not identify any studies that demonstrated improvement in patient activation in SCD, and we found no studies of patient portals as the eHealth intervention. Expanding the range of outcomes measured in the use of eHealth for self-management of SCD is a potential area for future research.

While we included articles that reported RCTs, quasi-experimental studies, or pilot pre-post studies, many preliminary studies and clinical trials are underway to develop and evaluate the next set of eHealth tools. These studies include preliminary needs assessments [17,80-83], processes for development of a tool [83,84], prototypes [84-87], pilot feasibility studies [31,83,88,89], and ongoing clinical trials [90,91] for eHealth interventions. A reason there may be fewer interventions published about SCD could be related to health information technology disparities with other diseases such as cystic fibrosis [92]. There is promise that mobile health technologies can help bridge this digital divide. With the increased uptake of mobile technology use and the number of preliminary studies in SCD, this is a prime area for future research. In addition, many studies have discussed improvement in self-management using eHealth in low- and middle-income countries [42-49]. While our review of the literature saw a paucity of studies from these countries, there is significant potential for eHealth to improve self-management of SCD in Africa, where the burden of SCD is much greater than in higher-income countries and the improvement in self-management with eHealth has been demonstrated [42-49]. Despite the promise of bridging the digital divide, lack of access to the mobile data plans required to deliver eHealth interventions could be a potential barrier to the wide dissemination of such interventions in middle- and low-income countries.

Despite the potential of eHealth to improve self-management of SCD, the SCD community and their health care providers need to exercise caution when using eHealth interventions. Many eHealth apps are available in online stores, but their

evaluation is lacking. As seen in a systematic review of pain apps, little research of the many apps available has been published [93]. Most apps have not been studied and are not regulated by governmental bodies. Use of these apps can come with significant risk to the accuracy of information delivered, as well as data privacy and security risks. The accuracy of the information included in different health apps is another major risk for users, and the associated costs with purchasing these apps could be a burden for many patients and a potential barrier to uptake. Evaluation of eHealth interventions as they are made available will be crucial in aiding providers and patients to choose eHealth tools that will be safe and effective in improving the care of people with chronic diseases.

Strengths and Limitations

Our systematic review has a number of strengths. First, in our review, we followed the recommendations for rigorous systematic review methodology [50,51,94-96]. Second, we conducted a review with a highly sensitive search strategy, guided by a medical information specialist, with no language restrictions so as to minimize publication bias and identify the largest possible number of relevant studies. Our search also included published systematic reviews, clinical trial registries, and various electronic databases. Third, although our search was limited to studies published since 1995, we identified no eligible studies before 2005, and therefore we believe that the possibility of missing earlier studies is very small. Fourth, 2 authors completed the review process independently at all stages of the systematic review.

Some potential methodological limitations of our systematic review warrant discussion. First, similar to any other systematic literature review, although we planned our search criteria to be as comprehensive as possible, the possibility of missing a few relevant articles cannot be excluded. Second, to identify the strongest available evidence, we included articles that were published in peer-reviewed journals, and therefore there could be a publication bias with the tendency to report positive study results [97]. Third, the study sample sizes and ages, and the definition of adherence to preventive behaviors and other related outcomes varied. These limitations prohibited a meta-analysis from being performed [98]. Fourth, some of the included studies had relatively small sample sizes.

Conclusions

Our systematic review is, to the best of our knowledge, the first to evaluate eHealth for self-management in pediatric and adult patients with SCD. We identified several eHealth interventions measuring a variety of outcomes, which showed improvement in multiple components of self-management of SCD. Despite the promising feasibility and acceptability of eHealth interventions in improving self-management of SCD, the evidence overall is modest. However, with the increased access to mobile technology, eHealth interventions have great potential to improve health outcomes in patients with SCD, as well as other chronic diseases. Future eHealth intervention studies are needed to evaluate their efficacy, effectiveness, and cost effectiveness in promoting self-management in patients with SCD using rigorous methods and theoretical frameworks with clearly defined clinical outcomes. This review describes the promise of eHealth to improve self-management in individuals with SCD; however, there are many areas of future research that can help demonstrate their usefulness in this population.

Acknowledgments

SMB was supported by grant number K12HS023011 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. RMC was supported by the National Heart, Lung, And Blood Institute of the US National Institutes of Health under award number K23HL141447. The authors also thank Ms Linda O'Dwyer (Galter Health Sciences Library, Northwestern University Feinberg School of Medicine, Chicago, IL, United States) for her support with the literature search.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies.

[PDF File (Adobe PDF File), 41KB - jmir_v20i7e10940_app1.pdf]

Multimedia Appendix 2

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.

[PDF File (Adobe PDF File), 66KB - jmir_v20i7e10940_app2.pdf]

Multimedia Appendix 3

Summary of included studies focused on eHealth interventions.

[PDF File (Adobe PDF File), 37KB - jmir_v20i7e10940_app3.pdf]

Multimedia Appendix 4

Summary of the components of eHealth interventions for all included studies.

[PDF File (Adobe PDF File), 261KB - jmir_v20i7e10940_app4.pdf]

References

1. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Dewi M, et al. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. Lancet 2013 Jan 12;381(9861):142-151 [FREE Full text] [doi: 10.1016/S0140-6736(12)61229-X] [Medline: 23103089]
2. Brousseau DC, Panepinto JA, Nimmer M, Hoffmann RG. The number of people with sickle-cell disease in the United States: national and state estimates. Am J Hematol 2010 Jan;85(1):77-78 [FREE Full text] [doi: 10.1002/ajh.21570] [Medline: 20029951]
3. Mvundura M, Amendah D, Kavanagh PL, Sprinz PG, Grosse SD. Health care utilization and expenditures for privately and publicly insured children with sickle cell disease in the United States. Pediatr Blood Cancer 2009 Oct;53(4):642-646. [doi: 10.1002/pbc.22069] [Medline: 19492318]
4. Steiner CA, Miller JL. Sickle cell disease patients in U.S. hospitals, 2004: Statistical Brief #21. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville, MD: Agency for Healthcare Research and Quality; 2006.
5. Yusuf HR, Atrash HK, Grosse SD, Parker CS, Grant AM. Emergency department visits made by patients with sickle cell disease: a descriptive study, 1999-2007. Am J Prev Med 2010 Apr;38(4 Suppl):S536-S541 [FREE Full text] [doi: 10.1016/j.amepre.2010.01.001] [Medline: 20331955]
6. Quinn CT, Rogers ZR, McCavit TL, Buchanan GR. Improved survival of children and adolescents with sickle cell disease. Blood 2010 Apr 29;115(17):3447-3452 [FREE Full text] [doi: 10.1182/blood-2009-07-233700] [Medline: 20194891]
7. Panepinto JA, Bonner M. Health-related quality of life in sickle cell disease: past, present, and future. Pediatr Blood Cancer 2012 Aug;59(2):377-385. [doi: 10.1002/pbc.24176] [Medline: 22522407]
8. Badawy SM, Thompson AA, Lai J, Penedo FJ, Rychlik K, Liem RI. Health-related quality of life and adherence to hydroxyurea in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2017 Jun;64(6):e26369. [doi: 10.1002/pbc.26369] [Medline: 27896936]
9. Lanzkron S, Carroll CP, Haywood C. The burden of emergency department use for sickle-cell disease: an analysis of the national emergency department sample database. Am J Hematol 2010 Oct;85(10):797-799 [FREE Full text] [doi: 10.1002/ajh.21807] [Medline: 20730795]
10. Centers for Disease Control and Prevention. Sickle cell disease: data & statistics. Atlanta, GA: CDC; 2016. URL: https://www.cdc.gov/ncbddd/sicklecell/data.html [accessed 2018-07-11] [WebCite Cache ID 70pj2sRWi]
11. Wagner JL, Alberts SR, Sloan JA, Cha S, Killian J, O'Connell MJ, et al. Incremental costs of enrolling cancer patients in clinical trials: a population-based study. J Natl Cancer Inst 1999 May 19;91(10):847-853. [Medline: 10340904]
12. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [Medline: 12401421]
13. Badawy SM, Thompson AA, Lai J, Penedo FJ, Rychlik K, Liem RI. Adherence to hydroxyurea, health-related quality of life domains, and patients' perceptions of sickle cell disease and hydroxyurea: a cross-sectional study in adolescents and young adults. Health Qual Life Outcomes 2017 Jul 05;15(1):136 [FREE Full text] [doi: 10.1186/s12955-017-0713-x] [Medline: 28679417]
14. Badawy SM, Thompson AA, Penedo FJ, Lai J, Rychlik K, Liem RI. Barriers to hydroxyurea adherence and health-related quality of life in adolescents and young adults with sickle cell disease. Eur J Haematol 2017 Mar 17;98(6):608-614. [doi: 10.1111/ejh.12878] [Medline: 28306171]
15. Badawy SM, Thompson AA, Liem RI. Beliefs about hydroxyurea in youth with sickle cell disease. Hematol Oncol Stem Cell Ther 2018 Jan 31 [FREE Full text] [doi: 10.1016/j.hemonc.2018.01.001] [Medline: 29397333]
16. Utrankar A, Mayo-Gamble TL, Allen W, Novak L, Kassim AA, Bonnet K, et al. Technology use and preferences to support clinical practice guideline awareness and adherence in individuals with sickle cell disease. J Am Med Inform Assoc 2018 May 07:doi: 10.1093/jamia/ocy036-doi:[Epub ahead of print]. [doi: 10.1093/jamia/ocy036] [Medline: 29741695]
17. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
18. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
19. Fox S. Health information online. Washington, DC: Pew Research Center; 2014. URL: http://www.pewinternet.org/2014/02/13/health-information-online-2/ [accessed 2018-07-11] [WebCite Cache ID 70pjQRC9V]
20. Mobile fact sheet. Washington, DC: Pew Research Center; 2018 Feb 05. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-07-11] [WebCite Cache ID 70pjaWgOs]
21. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]

22. 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]
23. Peters AE, Keeley EC. Patient engagement following acute myocardial infarction and its influence on outcomes. Am J Cardiol 2017 Nov 01;120(9):1467-1471. [doi: 10.1016/j.amjcard.2017.07.037] [Medline: 28843395]
24. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]
25. Haas K, Martin A, Park KT. Text message intervention (TEACH) improves quality of life and patient activation in celiac disease: a randomized clinical trial. J Pediatr 2017 Jun;185:62-67.e2 [FREE Full text] [doi: 10.1016/j.jpeds.2017.02.062] [Medline: 28343658]
26. John ME, Samson-Akpan PE, Etowa JB, Akpabio II, John EE. Enhancing self-care, adjustment and engagement through mobile phones in youth with HIV. Int Nurs Rev 2016 Dec;63(4):555-561. [doi: 10.1111/inr.12313] [Medline: 27653013]
27. Knoerl R, Lee D, Yang J, Bridges C, Kanzawa-Lee G, Lita SG, et al. Examining the impact of a web-based intervention to promote patient activation in chemotherapy-induced peripheral neuropathy assessment and management. J Cancer Educ 2017 Mar 06:doi: 10.1007/s13187-017-1200-0-doi: 10.1[Epub ahead of print]. [doi: 10.1007/s13187-017-1200-0] [Medline: 28265863]
28. Milani RV, Lavie CJ, Bober RM, Milani AR, Ventura HO. Improving hypertension control and patient engagement using digital tools. Am J Med 2017 Jan;130(1):14-20. [doi: 10.1016/j.amjmed.2016.07.029] [Medline: 27591179]
29. Solomon M, Wagner SL, Goes J. Effects of a Web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012;14(1):e32 [FREE Full text] [doi: 10.2196/jmir.1924] [Medline: 22353433]
30. Cronin RM, Hankins JS, Adams-Graves P, Thompson AA, Kalinyak K, Byrd J, et al. Barriers and facilitators to research participation among adults, and parents of children with sickle cell disease: a trans-regional survey. Am J Hematol 2016 Dec;91(10):E461-E462 [FREE Full text] [doi: 10.1002/ajh.24483] [Medline: 27421796]
31. Shah N, Jonassaint J, De Castro L. Patients welcome the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2014;38(2):99-103. [doi: 10.3109/03630269.2014.880716] [Medline: 24512633]
32. Al-Durra M, Torio M, Cafazzo JA. The use of behavior change theory in Internet-based asthma self-management interventions: a systematic review. J Med Internet Res 2015;17(4):e89 [FREE Full text] [doi: 10.2196/jmir.4110] [Medline: 25835564]
33. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012;12:CD007459. [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
34. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]
35. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
36. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
37. 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]
38. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
39. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
40. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
41. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
42. Abaza H, Marschollek M. mHealth application areas and technology combinations. a comparison of literature from high and low/middle income countries. Methods Inf Med 2017 Aug 08;56(7):e105-e122. [doi: 10.3414/ME17-05-0003] [Medline: 28925418]

43. Cho Y, Lee S, Islam SMS, Kim S. Theories applied to m-health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018 Feb 13. [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
44. Feroz A, Perveen S, Aftab W. Role of mHealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. BMC Health Serv Res 2017 Nov 07;17(1):704 [FREE Full text] [doi: 10.1186/s12913-017-2664-7] [Medline: 29115992]
45. Hurt K, Walker RJ, Campbell JA, Egede LE. mHealth interventions in low and middle-income countries: a systematic review. Glob J Health Sci 2016 Sep 01;8(9):54429. [doi: 10.5539/gjhs.v8n9p183] [Medline: 27157176]
46. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, 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]
47. Oliver-Williams C, Brown E, Devereux S, Fairhead C, Holeman I. Using mobile phones to improve vaccination uptake in 21 low- and middle-income countries: systematic review. JMIR Mhealth Uhealth 2017 Oct 04;5(10):e148 [FREE Full text] [doi: 10.2196/mhealth.7792] [Medline: 28978495]
48. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
49. Ridgeway K, Dulli LS, Murray KR, Silverstein H, Dal SL, Olsen P, et al. Interventions to improve antiretroviral therapy adherence among adolescents in low- and middle-income countries: a systematic review of the literature. PLoS One 2018;13(1):e0189770 [FREE Full text] [doi: 10.1371/journal.pone.0189770] [Medline: 29293523]
50. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
51. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A. The Oxford Levels of Evidence 2. Oxford, UK: Centre for Evidence-Based Medicine; 2011. URL: http://www.cebm.net/index.aspx?o=5653 [accessed 2018-07-11] [WebCite Cache ID 70pn8NHxT]
52. Bakshi N, Smith ME, Ross D, Krishnamurti L. Novel metrics in the longitudinal evaluation of pain data in sickle cell disease. Clin J Pain 2017 Jun;33(6):517-527. [doi: 10.1097/AJP.0000000000000431] [Medline: 27584817]
53. Creary SE, Gladwin MT, Byrne M, Hildesheim M, Krishnamurti L. A pilot study of electronic directly observed therapy to improve hydroxyurea adherence in pediatric patients with sickle-cell disease. Pediatr Blood Cancer 2014 Jun;61(6):1068-1073. [doi: 10.1002/pbc.24931] [Medline: 24436121]
54. Estepp JH, Winter B, Johnson M, Smeltzer MP, Howard SC, Hankins JS. Improved hydroxyurea effect with the use of text messaging in children with sickle cell anemia. Pediatr Blood Cancer 2014 Nov;61(11):2031-2036. [doi: 10.1002/pbc.25177] [Medline: 25132074]
55. Ezenwa MO, Yao Y, Engeland CG, Molokie RE, Wang ZJ, Suarez ML, et al. A randomized controlled pilot study feasibility of a tablet-based guided audio-visual relaxation intervention for reducing stress and pain in adults with sickle cell disease. J Adv Nurs 2016 Jun;72(6):1452-1463. [doi: 10.1111/jan.12895] [Medline: 26768753]
56. Gallo AM, Wilkie DJ, Wang E, Labotka RJ, Molokie RE, Stahl C, et al. Evaluation of the SCKnowIQ Tool and Reproductive CHOICES Intervention Among Young Adults With Sickle Cell Disease or Sickle Cell Trait. Clin Nurs Res 2014 Aug;23(4):421-441. [doi: 10.1177/1054773813479377] [Medline: 23572406]
57. Hardy SJ, Hardy KK, Schatz JC, Thompson AL, Meier ER. Feasibility of home-based computerized working memory training with children and adolescents with sickle cell disease. Pediatr Blood Cancer 2016 Sep;63(9):1578-1585. [doi: 10.1002/pbc.26019] [Medline: 27227457]
58. Inoue S, Kodjebacheva G, Scherrer T, Rice G, Grigorian M, Blankenship J, et al. Adherence to hydroxyurea medication by children with sickle cell disease (SCD) using an electronic device: a feasibility study. Int J Hematol 2016 Aug;104(2):200-207. [doi: 10.1007/s12185-016-2027-x] [Medline: 27225236]
59. Jacob E, Duran J, Stinson J, Lewis MA, Zeltzer L. Remote monitoring of pain and symptoms using wireless technology in children and adolescents with sickle cell disease. J Am Assoc Nurse Pract 2013 Jan;25(1):42-54 [FREE Full text] [doi: 10.1111/j.1745-7599.2012.00754.x] [Medline: 23279278]
60. Jonassaint CR, Shah N, Jonassaint J, De Castro L. Usability and feasibility of an mHealth intervention for monitoring and managing pain symptoms in sickle cell disease: the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2015;39(3):162-168. [doi: 10.3109/03630269.2015.1025141] [Medline: 25831427]
61. Leonard S, Anderson LM, Jonassaint J, Jonassaint C, Shah N. Utilizing a novel mobile health "selfie" application to improve compliance to iron chelation in pediatric patients receiving chronic transfusions. J Pediatr Hematol Oncol 2017 Apr;39(3):223-229. [doi: 10.1097/MPH.0000000000000743] [Medline: 28099398]
62. McClellan CB, Schatz JC, Puffer E, Sanchez CE, Stancil MT, Roberts CW. Use of handheld wireless technology for a home-based sickle cell pain management protocol. J Pediatr Psychol 2009 Jun;34(5):564-573 [FREE Full text] [doi: 10.1093/jpepsy/jsn121] [Medline: 19029141]
63. Pernell BM, DeBaun MR, Becker K, Rodeghier M, Bryant V, Cronin RM. Improving medication adherence with two-way short message service reminders in sickle cell disease and asthma. A feasibility randomized controlled trial. Appl Clin Inform 2017 May 24;8(2):541-559. [doi: 10.4338/ACI-2016-12-RA-0203] [Medline: 28536723]

64. Schatz J, Schlenz AM, McClellan CB, Puffer ES, Hardy S, Pfeiffer M, et al. Changes in coping, pain, and activity after cognitive-behavioral training: a randomized clinical trial for pediatric sickle cell disease using smartphones. Clin J Pain 2015 Jun;31(6):536-547. [doi: 10.1097/AJP.0000000000000183] [Medline: 25503599]
65. Wilkie DJ, Gallo AM, Yao Y, Molokie RE, Stahl C, Hershberger PE, et al. Reproductive health choices for young adults with sickle cell disease or trait: randomized controlled trial immediate posttest effects. Nurs Res 2013 Oct;62(5):352-361 [FREE Full text] [doi: 10.1097/NNR.0b013e3182a0316b] [Medline: 23995469]
66. Yoon SL, Godwin A. Enhancing self-management in children with sickle cell disease through playing a CD-ROM educational game: a pilot study. Pediatr Nurs 2007;33(1):60-3, 72. [Medline: 17411004]
67. Hazzard A, Celano M, Collins M, Markov Y. Effects of STARBRIGHT World on knowledge, social support, and coping in hospitalized children with sickle cell disease and asthma. Child Health Care 2002;31(1):69-86. [doi: 10.1207/S15326888CHC3101_5]
68. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Codina-Jane C. Does mHealth increase adherence to medication? Results of a systematic review. Int J Clin Pract 2015 Jan;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]
69. DeKoekkoek T, Given B, Given CW, Ridenour K, Schueller M, Spoelstra SL. mHealth SMS text messaging interventions and to promote medication adherence: an integrative review. J Clin Nurs 2015 Oct;24(19-20):2722-2735. [doi: 10.1111/jocn.12918] [Medline: 26216256]
70. Linn AJ, Vervloet M, van Dijk L, Smit EG, Van Weert JCM. Effects of eHealth interventions on medication adherence: a systematic review of the literature. J Med Internet Res 2011;13(4):e103 [FREE Full text] [doi: 10.2196/jmir.1738] [Medline: 22138112]
71. Rootes-Murdy K, Glazer KL, Van wert MJ, Mondimore FM, Zandi PP. Mobile technology for medication adherence in people with mood disorders: a systematic review. J Affect Disord 2018 Dec;227:613-617. [doi: 10.1016/j.jad.2017.11.022] [Medline: 29172054]
72. Tran N, Coffman JM, Sumino K, Cabana MD. Patient reminder systems and asthma medication adherence: a systematic review. J Asthma 2014 Jun;51(5):536-543. [doi: 10.3109/02770903.2014.888572] [Medline: 24506699]
73. Vervloet M, Linn AJ, van Weert JCM, de Bakker DH, Bouvy ML, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
74. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179 [FREE Full text] [doi: 10.2196/jmir.7290] [Medline: 28550009]
75. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
76. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013;12:CD007458. [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
77. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
78. Modi AC, Crosby LE, Hines J, Drotar D, Mitchell MJ. Feasibility of web-based technology to assess adherence to clinic appointments in youth with sickle cell disease. J Pediatr Hematol Oncol 2012 Apr;34(3):e93-e96 [FREE Full text] [doi: 10.1097/MPH.0b013e318240d531] [Medline: 22278205]
79. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
80. Anderson C, Torres S, Gokuli S, Brandow AM, Panepinto JA. Assessment of communication modes in patients and families with sickle cell disease. Pediatr Blood Cancer 2013 May;60(5):887. [doi: 10.1002/pbc.24477] [Medline: 23450782]
81. Issom D, Hartvigsen G, Bonacina S, Koch S, Lovis C. User-centric eHealth tool to address the psychosocial effects of sickle cell disease. Stud Health Technol Inform 2016;225:627-628. [Medline: 27332283]
82. Vaughn LM, McLinden D, Jacquez F, Crosby L, Slater S, Mitchell M. Understanding the social networks of parents of children with sickle cell disease. J Health Care Poor Underserved 2011 Aug;22(3):1014-1029 [FREE Full text] [doi: 10.1353/hpu.2011.0087] [Medline: 21841293]
83. Crosby LE, Ware RE, Goldstein A, Walton A, Joffe NE, Vogel C, et al. Development and evaluation of iManage: a self-management app co-designed by adolescents with sickle cell disease. Pediatr Blood Cancer 2017 Jan;64(1):139-145. [doi: 10.1002/pbc.26177] [Medline: 27574031]
84. Craven M, Lang A, Martin J. Developing mHealth apps with researchers: multi-stakeholder design considerations. In: Marcus 1, editor. Design, User Experience, and Usability: User Experience Design for Everyday Life Applications and Services: Third international Conference, DUXU 2014. Cham, Switzerland: Springer; 2014:15-24.

85. Cheng C, Brown C, New T, Stokes TH, Dampier C, Wang MD. SickleREMOTE: a two-way text messaging system for pediatric sickle cell disease patients. 2012 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics; Jan 2-7, 2012; Hong Kong and Shenzhen, China p. 408-411.
86. Washington G, Ward J, Kameka M. Spare me: towards an empathetic tool for helping adolescents & teenagers cope with sickle cell. 2015 Presented at: ICHI 2015: IEEE International Conference on Healthcare Informatics; Oct 21-23, 2015; Dallas, TX, USA p. 555-561.
87. Durfee R, Venugopalan J, Ren J, Wang M. Med-vest: a wearable sensory platform. 2014 Presented at: IEEE Healthcare Innovation Conference (HIC); Oct 8-10, 2014; Seattle, WA, USA.
88. Cheng C, Brown RC, Cohen LL, Venugopalan J, Stokes TH, Wang MD. iACT--an interactive mHealth monitoring system to enhance psychotherapy for adolescents with sickle cell disease. Conf Proc IEEE Eng Med Biol Soc 2013;2013:2279-2282. [doi: 10.1109/EMBC.2013.6609992] [Medline: 24110179]
89. Yang S, Jacob E, Gerla M. Web-based mobile e-diary for youth with sickle cell disease. 2012 Presented at: IEEE Consumer Communications and Networking Conference (CCNC); Jan 14-17, 2012; Las Vegas, NV, USA p. 385-389.
90. Creary S, Chisolm DJ, O'Brien SH. ENHANCE-(Electronic Hydroxyurea Adherence): a protocol to increase hydroxyurea adherence in patients with sickle cell disease. JMIR Res Protoc 2016 Oct 03;5(4):e193 [FREE Full text] [doi: 10.2196/resprot.6403] [Medline: 27697749]
91. Makubi A, Sasi P, Ngaeje M, Novelli EM, Mmbando BP, Gladwin MT, et al. Rationale and design of mDOT-HuA study: a randomized trial to assess the effect of mobile-directly observed therapy on adherence to hydroxyurea in adults with sickle cell anemia in Tanzania. BMC Med Res Methodol 2016 Dec 18;16(1):140 [FREE Full text] [doi: 10.1186/s12874-016-0245-9] [Medline: 27756209]
92. Barriteau CM, Thompson AL, Meier ER, Pecker LH. Sickle cell disease related internet activity is three times less frequent than cystic fibrosis related internet activity. Pediatr Blood Cancer 2016 Nov;63(11):2061-2062. [doi: 10.1002/pbc.26122] [Medline: 27362449]
93. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
94. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Oxford, UK: The Cochrane Collaboration; 2011. URL: http://handbook-5-1.cochrane.org/ [accessed 2018-07-11] [WebCite Cache ID 70poh61bf]
95. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. York, UK: University of York; 2009. URL: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf [accessed 2018-07-11] [WebCite Cache ID 70poZLTUL]
96. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
97. Dubben H, Beck-Bornholdt H. Systematic review of publication bias in studies on publication bias. BMJ 2005 Aug 20;331(7514):433-434 [FREE Full text] [doi: 10.1136/bmj.38478.497164.F7] [Medline: 15937056]
98. Bartolucci AA, Hillegass WB. Overview, strengths, and limitations of systematic reviews and meta-analyses. In: Chiappelli F, Caldeira XM, Neagos N, Oluwadara OO, Ramchandani MH, editors. Evidence-Based Practice: Toward Optimizing Clinical Outcomes. Berlin, Germany: Springer-Verlag; 2010:17-34.

Abbreviations

eHealth: electronic health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SCD: sickle cell disease

Edited by G Eysenbach; submitted 02.05.18; peer-reviewed by SE Brown, S Creary; comments to author 24.05.18; revised version received 06.06.18; accepted 21.06.18; published 19.07.18.

Please cite as:
Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, Shah N
Patient-Centered eHealth Interventions for Children, Adolescents, and Adults With Sickle Cell Disease: Systematic Review
J Med Internet Res 2018;20(7):e10940
URL: http://www.jmir.org/2018/7/e10940/
doi:10.2196/10940
PMID:30026178

Policy Proposal

Reimagining Health Data Exchange: An Application Programming Interface–Enabled Roadmap for India

Satchit Balsari[1,2], MD, MPH; Alexander Fortenko[3], MD, MPH; Joaquín A Blaya[4], PhD; Adrian Gropper[5], MD; Malavika Jayaram[6], LLM; Rahul Matthan[7], LLM; Ram Sahasranam[8]; Mark Shankar[9], MD; Suptendra N Sarbadhikari[10], PhD; Barbara E Bierer[11], MD; Kenneth D Mandl[12,13], MD; Sanjay Mehendale[14], MD, MPH; Tarun Khanna[9], PhD

[1]Beth Israel Deaconess Medical Center, Harvard Medical School, Department of Emergency Medicine, Boston, MA, United States
[2]Harvard FXB Center for Health and Human Rights, Boston, MA, United States
[3]NewYork-Presyterian Hospital, Emergency Medicine, New York, NY, United States
[4]The Human Diagnosis Project, Washington DC, DC, United States
[5]Patient Privacy Rights, Boston, MA, United States
[6]Digital Asia Hub, Hong Kong, China
[7]TriLegal, Bangalore, India
[8]Athenahealth, San Francisco, CA, United States
[9]Harvard Business School, Boston, MA, United States
[10]International Institute of Health Management Research, New Delhi, India
[11]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States
[12]Boston Children's Hospital, Computational Health Informatics Program, Boston, MA, United States
[13]Harvard Medical School, Department of Pediatrics and Department of Biomedical Informatics, Boston, MA, United States
[14]Indian Council of Medical Research, New Delhi, India

Corresponding Author:
Satchit Balsari, MD, MPH
Harvard FXB Center for Health and Human Rights
651 Huntington Avenue, 703C
Boston, MA,
United States
Phone: 1 6174320011
Email: sbalsari@bidmc.harvard.edu

Abstract

In February 2018, the Government of India announced a massive public health insurance scheme extending coverage to 500 million citizens, in effect making it the world's largest insurance program. To meet this target, the government will rely on technology to effectively scale services, monitor quality, and ensure accountability. While India has seen great strides in informational technology development and outsourcing, cellular phone penetration, cloud computing, and financial technology, the digital health ecosystem is in its nascent stages and has been waiting for a catalyst to seed the system. This National Health Protection Scheme is expected to provide just this impetus for widespread adoption. However, health data in India are mostly not digitized. In the few instances that they are, the data are not standardized, not interoperable, and not readily accessible to clinicians, researchers, or policymakers. While such barriers to easy health information exchange are hardly unique to India, the greenfield nature of India's digital health infrastructure presents an excellent opportunity to avoid the pitfalls of complex, restrictive, digital health systems that have evolved elsewhere. We propose here a federated, patient-centric, application programming interface (API)–enabled health information ecosystem that leverages India's near-universal mobile phone penetration, universal availability of unique ID systems, and evolving privacy and data protection laws. It builds on global best practices and promotes the adoption of human-centered design principles, data minimization, and open standard APIs. The recommendations are the result of 18 months of deliberations with multiple stakeholders in India and the United States, including from academia, industry, and government.

(J Med Internet Res 2018;20(7):e10725) doi:10.2196/10725

KEYWORDS

health information exchange; India; health APIs

Introduction

Background

India's population of over 1.3 billion is served by over 2.5 million health care workers of varying qualifications. The vast majority of clinical interactions are not digitized. In the few instances that they are, the data are not standardized, not interoperable, and not readily accessible to clinicians, researchers, or policy makers [1]. While barriers to easy health information exchange (HIE) are hardly unique to India, the greenfield nature of India's digital health infrastructure presents an excellent opportunity to avoid the pitfalls of complex, restrictive, digital health systems that have evolved elsewhere.

In February 2018, the Government of India announced a massive public health insurance program under the National Health Protection Scheme (NHPS), offering Indian Rs 500,000 (approximately US $ 7,600) in annual coverage to 100 million households, or nearly 500 million citizens [2]. To meet this bold target, the government will rely on technology to effectively scale services, monitor quality, and ensure accountability. While India has seen great strides in informational technology development and outsourcing, mobile phone penetration, cloud computing, and financial technology, the digital health ecosystem is in its nascent stages and has been waiting for a catalyst to seed the system. The NHPS is expected to provide just this impetus for widespread adoption.

We propose here a federated, patient-centric, application programming interface (API)–enabled health information ecosystem that leverages India's near-universal mobile phone penetration, universal availability of unique identification (ID) systems, and evolving privacy and data protection laws. The arguments laid out here are the result of an extended set of deliberations that began at an interdisciplinary seminar held at Harvard in September 2016 and have since resulted in potential pathways for prototype development in India.

The State of Health Data Exchange

Electronic Health Records (EHRs) have traditionally been closed systems, sometimes incapable of sharing access across platforms within the same institution, and almost never across vendors at independent institutions. While more systems now allow patients access to their health-related data, few EHRs give patients control over how their data will move across institutions or be shared between providers. Despite significant legislation, a large portion of health data collected today remains inaccessible due to legitimate concerns over confidentiality and privacy, risk-averse hospital policies, prohibitive costs associated with change, and inertia [3].

While health data have been typically associated with information captured in EHRs, there is growing recognition that data are generated at multiple nodes along the delivery system. For example, at the pharmacist, at the stand-alone imaging facility, at the laboratory, at the general practitioner's office, at the hospital, at the insurance company, and now, even on one's wrist [4]. However, the lack of standardization among data storage systems makes it virtually impossible to combine and collate data from multiple sources, resulting in duplication, redundancy, wastage, and delays [5].

The concept of a personal health record (PHR) has long been floated as one potential solution to disjointed health care data [4,6,7]. A PHR relies on a patient-controlled repository where data may be accessed from multiple nodes within the system. Standalone PHRs mostly rely on the patient's drive and ability to input data [8,9]. Tethered (ie, connected) PHRs are patient-accessible components of electronic medical records linked to an institution or health system [8,9]. Still, there are drawbacks. PHRs seldom allow direct input from or access to entities outside the network. Neither PHR model allows for the development of third-party applications (app) on the patient's health data repository. Although there is interest from the consumer, widespread adoption of both has been hindered by concerns about data ownership, interoperability, security, and scalability [10-12]. The Ministry of Health and Family Welfare (MoHFW) in India has demonstrated an interest in developing a PHR-based system [13].

In recent years, additional individual and population health data have been generated by wellness gadgets (eg, Fitbit), Web-based diagnostic devices (eg, AliveCor), patient-facing apps (eg, Stanford Healthcare), provider-facing apps (eg, Practo), or researcher-facing apps (like Apple's Research Kit). These new apps and gadgets create additional silos of health data. In fact, of the 260,000 mHealth apps that existed on the last count, 90% were free—their financial viability predicated on their ability to monetize the data they collect [14]. In the United States (US), the 21st Century Cures Act (2016), mandated that "certified" health information technology (IT) products have APIs that allow health information to be accessed, exchanged, and used "without special effort." Standardization was not mandated, making interoperability difficult to implement [15]. In India, while the mHealth industry is booming and expected to grow exponentially, there are no legal provisions to regulate access to personal health data that flow in and out of these devices and apps, and sometimes across international borders [16].

The call for data integration, universal compatibility, and portability has come from many quarters. There is no shortage of standards, but few are universally applied. There are standards for nomenclature and terminology, structural and semantic standards, and open source technology platforms that promote secure health information exchange [17]. Entrepreneurs and provider networks have responded to this need for data portability, and the potential for monetizing vast amounts of data, by creating their own ecosystems where the patient is not the final arbiter of data flow [18-21].

There is now sufficient recognition that restricting health data access is detrimental to patient care, provider satisfaction, and health care costs [22]. Conversely, access to health data has been shown to benefit and empower patients [23]. Authorized access to the vast troves of accumulated digital data helps

accelerate medical research [24]. In March 2016, the US National Institutes of Health in collaboration with the Office of the National Coordinator for Health IT announced the launch of the Precision Medicine Initiative Sync for Science program. Based on existing community standards and specification efforts, this pilot program gave patients easy access to their data and allowed them to donate it to researchers [25] securely. The most significant challenges facing this initiative are individual and collective concern over data security and privacy, and hospitals' and practitioners' reluctance to promote the program.

Health Data in India

Health is a "state" subject in India, managed and funded by state governments, with part-funding from the Center (ie, the federal government in Delhi). Consequently, there is wide variation in quality of care within and among states [26,27]. On average, 70% of health care is delivered through the private sector, which is comprised of state-of-the-art tertiary facilities, nursing homes, polyclinics, general practitioners and a large workforce of health care providers with no medical qualifications [28]. The public health system is robust in population health interventions like vaccine delivery but struggles to provide quality primary care or specialized services at scale [29]. There are also stark differences in health care services in rural and urban India, with the majority of the medical workforce, and tertiary services gravitating to urban India [30].

Conversations about health information exchange in India must acknowledge these realities, as well as the near absence of digital health information in most clinical transactions. The private sector is mostly not digitized except for major diagnostic laboratory and radiology franchises, and some private hospital networks [31,32]. Hospital chains like the Apollo Group and Max Healthcare Group with advanced EHRs have reached Stage 6 of the Healthcare Information and Management Systems Society classification of EHR adoption [33]. However, EHRs are not yet portable across these institutions, and most systems continue to struggle with physician compliance [31,32,34].

In the public sector, data have been collected through various overlapping, regional or national mandates, or dictated by the needs of sponsoring philanthropic foundations [35-37]. The systems, where they exist, are District Health Information System 2 compatible, and mostly comprised of aggregate data. Longitudinal patient records are a recent (and still rare) phenomena. Select public hospitals have managed to digitize some components of the medical record. The governments of Rajasthan, Andhra Pradesh, Maharashtra, and Tamil Nadu have all sponsored EHR implementation to varying extents [38,39]. The Tamil Nadu system, for example, connects over 1,500 primary health centers, 267 secondary care hospitals, and 17 medical colleges. Sustained investment by the state government has been key to the program's success and scale [39]. The JJ Group of Hospitals in Mumbai, Maharashtra has logged over 20 million patient visits in the past decade, using the Amrita HIS platform [40]. India's premier public hospital in Delhi, AIIMs, uses a patient scheduling software aimed at reducing wait times. Several other state government sponsored tertiary hospitals use the e-Aushadhi supply chain management system [41,42]. However, these systems lack portability. They are also limited to government-run health facilities precluding residents from accessing data across state lines or from transporting data across public, and private health care facilities.

At the primary care level, community health workers and clinical staff log data in paper-based notebooks, tablets, excel sheets, and a variety of software applications that differ from state to state. The validity of much of these data is questionable [43,44]. Preliminary analysis from a study by our team at a primary care center in one state in India shows that there are over 3,000 discreet fields of data captured in paper-based and electronic forms, contributing data to 70 different databases, the majority of which are never accessed or used for real-time clinical or policy decision making. These repetitive data collection requirements result in large duplication, wastage, and divergence of limited human resources.

Health data are also captured at hundreds of research institutions across the country, in paper files, personal flash drives, hard drives, and sometimes on institutional servers. There is a general consensus among local researchers that much of these data are never analyzed, and they seldom change clinical practice or care delivery [45,46].

Data do not travel across jurisdictions. For example, the Indian Council of Medical Research, India's leading body for biomedical research has limited access to the data generated across its various collaborating institutions, and almost none to data generated in private institutions. Critical clinical data with significant individual and public health consequence, like information on compliance and antibiotic resistance, are not portable across institutions. Also, the government's Revised National TB Control Program cannot follow patients or monitor their care once they choose to seek treatment in the private sector. Even if private sector entities were willing to share data, there are no mechanisms to do so. The lack of interoperability of such critical data has a profoundly negative consequences on existing disease surveillance systems [47].

There is growing recognition among the public sector that all new digitization efforts must conform to prescribed standards. The government adopted Systematized Nomenclature of Medicine (SNOMED) and is making it available free of cost to health systems across India [48]. Organizations like HIMSS and the India Health Information Network are other key stakeholders. While the government intends to establish interoperability standards, the question of change management remains unaddressed. Who will bear the cost of these new systems, and of transitioning the older systems? What will be the institutional and individual incentives? How will the system be seeded, populated and sustained? In the United States, the Health Information Technology for Economic and Clinical Health (HITECH) Act provided incentives (and penalties) for Electronic Health Record (EHR) adoption, at the cost of US $ 34 billion in payments to doctors and hospitals to purchase and promote electronic health systems [49]. A 2017 study comparing hospitals that qualified for monetary incentives for implementing "meaningful use" of EHRs to those that did not qualify, attributed 8 percentage points in adoption growth to the incentives provided by HITECH [50]. India can barely afford such expenditures when health spending is at less than 2% of

its GDP [51]. What then would be the much-needed catalyst to stimulate widespread digitization?

Until 2017, the lack of a significant insurance player in the market precluded piggybacking EHR adoption on the billing requirements of payers and providers. Deliberations with critical stakeholders in India facilitated by the authors through 2016 and 2017 focused on the three nodes in the system that were most digitized, namely, laboratories, pharmacies and radiology reports. Still, mechanisms for change were not clear. The 2018 NHPS, with its urgent need for a technological backbone, changes all that—it has potential to finally develop this vision for universal HIE in India.

This approach is, however, not without its dangers. The EHRs in the United States have evolved as very effective billing instruments and provided medicolegal safeguards, but basic patient and provider needs like portability and access were an afterthought and required prohibitively expensive retrofitting. Systems work best, and compliance is highest, when the EHRs can be customized to local workflows, and when they can be modified or upgraded with relative ease and at low costs. Hospital mergers in the United States have resulted in near-uniform systems across vastly different enterprises, changes in which require universal consensus across the ecosystem and entail prohibitive fees charged by EHR corporations and additional re-training costs.

While the NHPS may indeed be the much-awaited catalyst for jumpstarting the digital health ecosystem in India, mandating a one-size-fits-all nationwide billing platform will do irreparable and costly damage—costs the Indian health care system cannot afford to bear [52].

An Application Programming Interface–Enabled Health Information Exchange for India

Between August 2013 and December 2016, India's MoHFW released a set of recommendations for electronic health records that outlined vital components of a standardized health care information ecosystem, and a common language for the organization of medical terminology and data [43,53]. The Ministry also instituted the National Digital Health Authority meant to "regulate, develop and deploy digital health along the continuum of care across India [54]." In December 2016, the government's Centre for Health Informatics released a Request for Proposals for the creation of an integrated health information platform (IHIP), where the exchange would be facilitated via a central storage repository. Low budget allocation and limited information on the proposed architecture precluded most IT service companies from bidding. The contract was finally awarded in January 2018, just weeks before the NHPS was announced [55]. The centralized IHIP as envisioned would be monolithic and prescriptive, risking poor user adoption, high physician burnout, and little meaningful access to the vast data it would generate, as seen in large hospital systems in the United States [52,56]. It risks not being able to leverage future IT developments.

The prototype outlined below argues against the use of a centralized repository of health data. Instead, we submit that the way forward must be an API–enabled, blockchain-based information network in which the personal health record underpins a system where free, real-time flow of data is predicated on consent and authorized access. India is uniquely positioned to build this ecosystem armed with a universal identity system, experience with digitization across multiple other industries, and a sophisticated domestic IT workforce.

We describe below the technical and legal basis for the design proposed.

Federated Architecture

Our proposed model calls for a federated architecture that acknowledges current and future health information flow; for example, between providers and patients, wearables and EHRs, consumers and pharmacies, physicians and laboratories, or institutions and payers. Collating all data in a national repository for 1.3 billion Indians will prove to be prohibitively expensive, redundant, and wasteful. It would also offer a single point of failure where security breaches would result in colossal data compromise. A federated system would allow data to sit at the source and be recalled on demand.

An API–enabled federated health data architecture would function on blockchain principles as an "open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way" [57,58]. Consider a PHR that could query all nodes in the network to receive periodic updates—from wearables, diary entries, pharmacists, doctors, hospitals, diagnostic labs, imaging facilities, and payers. It is possible to map out various permissible pathways through which the data can travel automatically while there may be others through which it cannot pass without the patient's consent (Figure 1).

An authorized physician—even a virtual "teledoc"—would be able to call for her patient's entire record, either through pre-authorization, real-time authentication, or waivers in case of emergencies [56]. Diagnostic laboratories should be permitted to send their reports to the patient's physician who requested the test but will need authorization from the patient to send it to any other doctor (such as one to whom the patient goes for a second opinion). Similarly, a public health agency, duly authorized, should be able to query select de-identified test results across all laboratories in a region of interest, to forecast, monitor, and respond to disease outbreaks [59]. Health system administrators should similarly have access to aggregate data for monitoring delivery, resource utilization and clinical outcomes. Physicians should be able to query their practice patterns. Third-party applications that are built off the patient's PHR, for example, alerting the patient to vaccine requirement before travel, or triggering reminders based on her medication list, would need the patient's permission to access data from her PHR (Figure 2).

Figure 1. Federated Health Information Exchange Schema: The personal health record (PHR) would access data from existing and novel sources, by preauthorization, waiver or legal mandate. EHR: electronic health record; MOH: Ministry of Health.

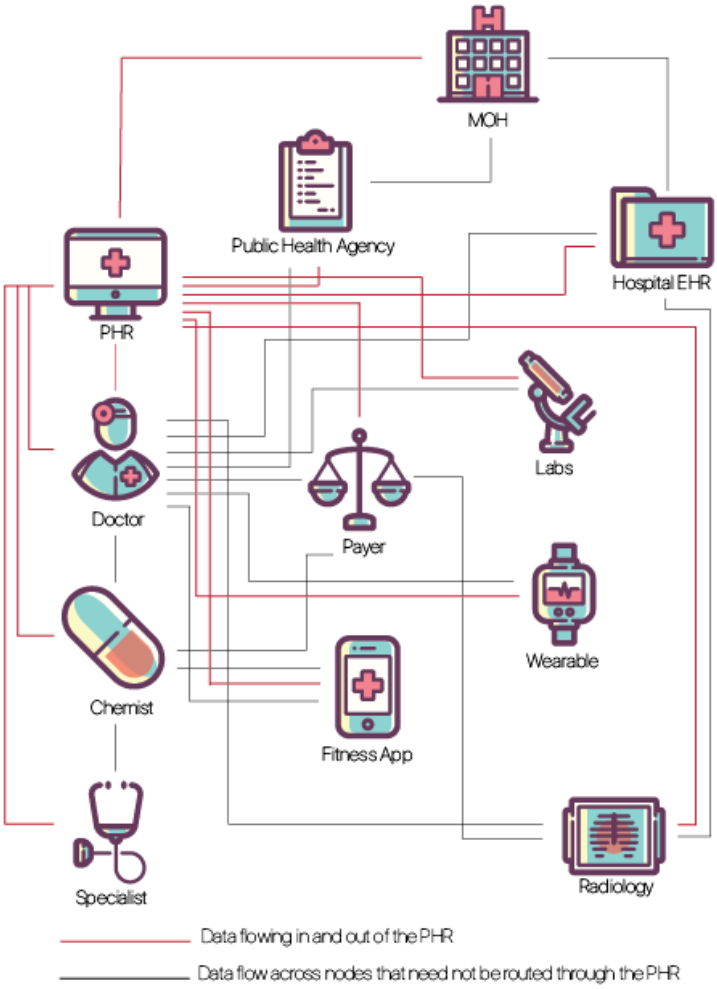
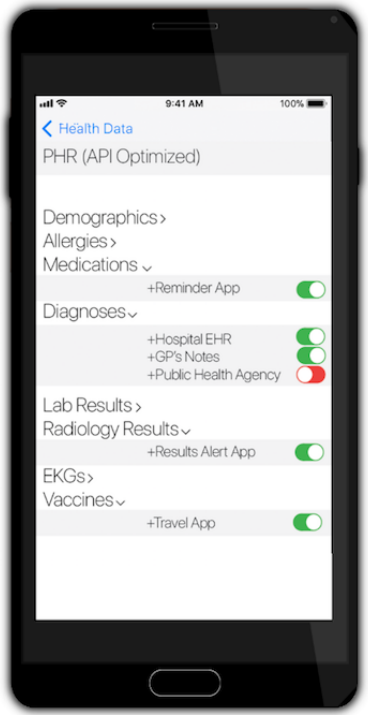


Figure 2. A customizable personal health record (PHR) interface. Through user-driven consent and control, third-party plug-ins (apps) can access the PHR via standard open application programming interfaces (APIs).



As long as the user is "recognized" by the system, and therefore has pre-authorization to query particular types of data, access should be easy and near instantaneous. Essentially, the "consent" process is separated in time and place from data flow, allowing timely, secure, exchange of relevant health data between nodes. A federated and distributed network so constructed would obviate the need for constructing large national or regional databases of the patient's "entire" medical record. Opportunistic synchronization and personal device-based back-up may reasonably mitigate the effect of unreliable electricity and connectivity. Moreover, all these inquiries and interactions should generate audit trails to prevent misuse.

Smartphone penetration in India is expected to reach about 36% of all mobile phone users [60]. While important components of the proposed architecture, mobile phones are not critical. Web-based services, hospital-based kiosks, and patient service facilities can assist those without mobile phones in understanding and accessing their settings. Local regulations should ensure that the default always protects the patient's interests and privacy.

While seemingly simple in architecture, such a system is predicated on standardization and widespread adoption. We next discuss how there is sufficient precedence and local capacity to favor such sweeping change.

Universal Unique Patient Identifier

Prima facie, access to the federated architecture would require a universal identifier. All data would be tagged with that unique identifier no matter where the patient interfaced with the medical system. Further, any entity contributing to or extracting from the system would need a unique institution and personal identity tag.

While usually a daunting system to create, the near universal penetration of India's unique biometric identification program, Aadhaar, offers a solution to this challenge. Aadhaar has been built around the principles of privacy by design, and data minimization, that are particularly relevant in security-sensitive applications like health care. Administered by the Unique Identification Authority of India, the system is actively used today for the central government's direct benefit transfers and subsidies programs and has also been used by several banks and telecom operators. By 2017 over a billion Indian residents were enrolled in Aadhaar, making it the most widely deployed single ID system anywhere in the world.

This astonishing penetration notwithstanding, mandating the linkage of public services to Aadhaar has been problematic, and a subject of constitutional challenges before the Supreme Court, with legitimate concerns of misuse and state overreach [61]. It is possible that the public resistance for using Aadhaar may necessitate the creation of a separate unique medical identity number.

Local Precedence

The use of APIs would underpin the proposed federated architecture. An API is a set of routines, protocols, and tools built into a software application that enables it to communicate easily with other applications. APIs provide the means to build interoperable software and data exchange services. APIs are already behind the seamless integration from which we benefit every day. They allow Instagram to access the camera on our phones, agnostic to the maker of the device; or allow WhatsApp to access our phone directory ("contact lists"); or a taxi app to use Google maps. Industries like banking, finance, and social media have already successfully tapped into the explosive growth of software applications by adopting API-based solutions [62,63]. Unrelated corporations own most products that communicate via APIs, yet the use of standard APIs powers the entire digital ecosystem.

India's own experience with wide-scale API adoption in the financial technology (eg, fin-tech) sector has been regarded as hugely successful: The Unified Payments Interface, rolled out in 2016, has demonstrated both the feasibility and the advantages of adopting an API-based ecosystem. Aadhar spurred a range of nationwide API-based IT solutions, collectively referred to as IndiaStack [64]. The Aadhar dashboard eKYC, for example, is a paperless "Know Your Customer" (KYC) process aimed at verifying an individual's identity. This instant electronic verification has replaced the traditional KYC process that relied on the onerous in person presentation of paper-based copies of official identity documents. The eKYC is currently used across many industries, including banking, utilities, and mobile services, and has logged 4.2 billion transactions in the past four years [65].

The Digital Locker is another successful application that provides a cloud-based storage service to all residents that authorized users can access. Registered Digital Locker organizations can push (or retrieve) electronic copies of documents and certificates (eg, driving license, voter ID, school certificates) directly into the lockers of Indian citizens, once again making credentialing and verification processes near-instantaneous. As of January 2018, nearly 2 billion digital documents have been issued through the Digital Locker API [66]. The widespread adoption of these changes in the financial-tech sector was catalyzed by support from the highest levels of government.

The successful and explosive use of mobile financial services notwithstanding, user-controlled dataflow through the federated network assumes some degree of digital literacy and understanding the ramifications of consent, secondary data use, artificial intelligence algorithms, and so on. Until such time that the Indian populace is assumed to have such knowledge, concomitant local laws will need to require that default data access protects foremost, the patient. We discuss the relevant existing and evolving Indian legal standards in subsequent sections.

Global Standards

Globally, the health care industry has begun embracing API-based exchange. Open Medical Record System, and platforms like SMART Health IT, developed at Boston Children's Hospital, have long pursued health data ecosystems anchored in open standard APIs [67,68]. These programs, while successful, had been limited in scale due to the rigid and expensive architecture discussed above. However, the field is rapidly changing. In June 2017, Athenahealth invested US $63

million in a product that improved user-experience and ensured interoperability with other EHRs. The product enabled Athenahealth to open up their APIs for accelerated innovation for their provider base [69]. In January 2018, Apple adopted SMART standards for their Health Kit; others are expected to follow suit. Patients will soon come to expect that their health data be accessed via a range of apps on their phone.

Successful interoperability will rely on widely adopted standardization in data storage and retrieval. Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT), Logical Observation Identifiers Names and Codes (LOINC), and RxNorm, a standardizing nomenclature for a medication that can interpret varying vocabulary used by pharmacy and drug interaction software, are increasingly used globally. Health Level 7 (HL7) enables health records and exchanges to be built with typical architecture and structure. Fast Healthcare Interoperability Resources (FHIR), building off HL7 standards, provides data formats and resources for building APIs for facilitating exchange. Project Argonaut, for example, a consortium involving governmental, private and academic health IT leaders, has incorporated these standards to begin work on supporting the uptake of APIs and including them in "meaningful use" regulation [17,70]. India was an early adopter of many of these standards but remains a slow implementer, given the absence of incentives (or penalties) for adoption.

Data Minimization

The architects of India's digital health infrastructure, while being compliant with global standards, may consider creating a series of "minimum datasets" for standardization and interoperability. In this article, the use of the term *minimum data set* denotes the least possible number of data points that early digital health information systems in India must include to be useful to a range of stakeholders. Obfuscating digitization with validity, new digitized data collection tools are frequently bloated with information that will either not be used or is already being collected elsewhere. A minimal viable product, instead, will help seed the ecosystem and can be incrementally expanded upon, not unlike the phased requirements seen with HITECH. This approach will allow a vast range of pilots to be tested, evaluated, optimized to meet contextually relevant needs, and then scaled up. Data minimization is also a useful tool in improving security and reducing privacy abuse [71]. In India, focusing on *structured* laboratory data, radiology results, medications lists (ie, in sync with interactions at the chemist), allergies, diagnoses, and essential demographic information would provide enough data to spurn a variety of applications for all involved stakeholders.

Substitutability

It is imperative that early prototypes pay critical attention to the user experience. Attempts at EHR adoption in India have failed to date due to the untenable combination of very high patient volume and poor usability. Once again, an API–enabled system will allow providers and institutions to select products that are highly customized to the local context and workflow [72]. Such "substitutability" will be central to advancement—apps that access underlying data that can be updated or swapped out for new ones with improved features and usability, just as we currently do with apps from an app store [15].

The Law

Traditional privacy principles have been articulated in the Organization for Economic Cooperation and Development guidelines first published in the 1980s, revised last in 2013, the European Directive 95/46/EC, and several pieces of national legislation. These principles have come under increasing scrutiny with the power of big data analytics to combine information from discreet datasets. The EU General Data Protection Regulation (GDPR), considered one of the most stringent of data protection laws came into effect on May 25, 2018. It adopts a rights-based framework, placing the individual at the center of the law. In the United States, while the public sector is mostly governed by the Fair Information Practice Principles and related acts, data flow in the private sector is primarily regulated by notice and consent and overseen mostly by the Federal Trade Commission.

As India's planners imagine its new technologically powered health data ecosystem, hard questions need to be answered. For example, what risks do we pose for individuals and populations by allowing such seamless data travel? Who owns the data? Can such data be sold? If yes, does the patient have a financial claim, even when data are de-identified? What protection measures need to be put in place? What remedy does the patient have? What legal risks do patients, providers, scientists, and governments expose themselves or each other to? Are the technologies for such secure, encrypted, failsafe ecosystems available? What can we learn from other industries? Why have previous attempts failed?

Emerging economies often lack dedicated privacy laws, relying instead on a patchwork of consumer protection laws, telecommunications statutes, human rights provisions and other measures to tackle data breaches, privacy violations, and constitutionally protected rights to equal treatment. However, as government welfare and benefits are increasingly delivered through online platforms on the backs of newly digitized databases, there is a need to ramp up the legal infrastructure in parallel [73]. This need is critically felt when examining the ability of illiterate users to provide informed consent and to exercise control over valuable data. By engineering interoperable systems that default to protect and empower users by offering them control and discretion over data sharing arrangements, one can optimize the benefits of exchange without compromising privacy or security.

Data mining of de-identified information can now reveal very sensitive data [74-76]. Insurance premiums (in a less regulated health care system), for example, can be modified based on zip codes, browser history or seemingly unrelated shopping habits [75,77]. Technology is not neutral, and most systems encode values and biases, however unconscious [78,79]. Evolving research highlights the risks of data-driven or algorithmic decision-making [80,81]. Biometrics, which systems rely upon for identity verification, has been shown to have higher error rates (false positives and false negatives) for darker skin tones [82].

Societal expectations of privacy can also ebb and flow, and the laws need to be nimble enough to accommodate for the fast pace of IT evolution. Only a few years ago, the idea that Google would scan emails to automatically populate a person's calendar or send alerts was considered highly unacceptable. Today, for many, it is the norm.

In August 2017, the Supreme Court of India ruled that privacy is a fundamental right [83]. The landmark judgment potentially provides the necessary deterrents to data misuse in a health IT ecosystem, in addition to establishing the supremacy of patient's control over her health data (privacy). The Ministry of Electronics and Information Technology constituted an expert committee to help draft a bill on data protection. The committee released a White Paper on the data protection framework for India, inviting public comment [84]. The White Paper extensively examined best practices elsewhere in the world, with particular attention to European Commission's GDPR regulations scheduled to go into effect on May 25, 2018.

Responses to the White Paper included recommendations focused on the protection and use of health data, calling for automated but consented flows, easier access, and portability, and without jeopardizing the safety or privacy of vast swathes of India's digitally illiterate populations. Reviewers opined that health data are generated jointly by the patient and the provider, and are used for purposes beyond clinical care, including for research, operations, payments, quality control, and public health. The patient serves as the "data controller" with a reasonable say in which of their data are made available, to whom, and when. A "data processor" co-creates and adds data to the patient's health record and accesses it when implicitly or explicitly authorized to do so. When patients cannot consent for lack of capacity, illiteracy, or circumstance, regulations should favor the patient's best interest [85]. It might be useful to think about control in the context of a tiered hierarchy of permissions. The patient in most respects would be the final arbiter. Below the patient is a category of stakeholders with access to the data because they have been involved in its creation and to whom the patient has given implicit or explicit consent. Subordinate to these creators of data will be various other actors who can only gain access to the data with the patient's permission or, if permitted by law, without [86].

In March 2018, the MoHFW invited public comment on a new bill it has proposed, the Digital Information Security in Health Care Act [87]. These evolving privacy and data protection laws in India must provide for deterrents to data misuse, where violations result in hefty fines, loss of access, and censure. On the other hand, an onerous, manual, consent-driven process would prevent patients from benefiting from advances in voice-activated services, machine learning, and artificial intelligence. Users of health data must therefore, above all, be expected to have a fiduciary responsibility toward the patient.

Conclusion

In July 2018, the Government of India's NITI Aayog, the National Institute for Transforming India, published a blueprint for a "National Health Stack," embracing the principles of federated patient-centric data flows outlined in this paper [88,89]. The proposed "Health Stack" platform has the potential for revolutionizing medical care, research, and health care delivery in India. However, a simple, claims-driven model risks replicating the pitfalls of the US system [55]. It is also not likely to inspire provider or patient adoption.

For this proposed technological framework to meet its game-changing potential, the model will benefit from adhering to the following principles: (1) adopting a federated architecture, (2) prioritizing patient and population health needs over billing needs, (3) guaranteeing a patient's right to her structured data, (4) allowing a plug and play model of highly customizable applications that can address varying context-specific needs, and that respond to market incentives for better user-interfaces, (5) mandating minimum data sets, (6) adopting privacy by design: automate audited and consented data flow, and finally (7) defaulting to safeguarding patients' control over their data.

For patients, scientists and clinical providers to recognize, adopt and benefit from the vast potential of a secure, federated health information ecosystem, we propose a suite of initial applications whose benefits to society are palpable. For example, medication alerts, laboratory trends, schedulers and payment logs would prove highly useful to patients but would require interoperability among different sources of data, mandated or incentivized by the state. At a population level, disease surveillance data for modeling and forecasting outbreaks would be particularly useful to public health agencies. Standardized registries for trauma, cancer, rare diseases are desperately needed in India and can be built on the proposed framework. Aggregated and anonymized data sets accessed through an audited trail would help accelerate medical research, given the sheer volume of patient load in India.

However, such widespread adoption and data transfer between entities would necessitate buy-in from multiple stakeholders—through a combination of incentivization, legal mandate, budgetary allocation, and market demand for patient–provider, provider–provider, provider–payer, and payer–patient interactions [22]. There are no PHRs in India. It is precisely the greenfield nature of the digital health ecosystem in India that would allow a PHR-based, API–enabled network from the get-go, pre-empting the development of complex and incompatible silos of health data.

India must take advantage of its vibrant IT ecosystem and the widespread adoption of mobile technologies across its socio-economic strata. A light and robust API–enabled spine upon which both the public and private sector can be invited to build contextually appropriate, competing, substitutable, and incremental solutions will be the key to a forward-looking digital health ecosystem. The time for large centralized data warehouses and homogenous systems has passed.

There is much excitement globally about the power of big data, artificial intelligence and machine learning in reshaping medicine and health care delivery. However, the potential of these promising sciences can only be harnessed with reliable and timely data. A federated PHR in India will provide unprecedented amounts of health data; among them may lie answers to our well-being and happiness.

Acknowledgments

The authors would like to acknowledge the support of the Radcliffe Institute for Advanced Studies at Harvard University for sponsoring the workshop, “Exploring health information exchange: setting an interdisciplinary agenda,” in September 2016, where the first draft of this paper was proposed.

Authors' Contributions

SB wrote the first and final drafts of this paper, with significant editorial contributions from AF. All other authors contributed equally.

Conflicts of Interest

None declared.

References

1. Kurian O. The Wire. Why It's a Challenge To Make Quick Sense of India's Health Data URL: https://thewire.in/health/imr-mmr-data-nss-srs [accessed 2018-03-25] [WebCite Cache ID 6yBX212wE]
2. Chatterjee P. National Health Protection Scheme revealed in India. The Lancet 2018 Feb;391(10120):523-524. [doi: 10.1016/S0140-6736(18)30241-1]
3. McGraw D, Dempsey JX, Harris L, Goldman J. Privacy as an enabler, not an impediment: building trust into health information exchange. Health Aff (Millwood) 2009;28(2):416-427 [FREE Full text] [doi: 10.1377/hlthaff.28.2.416] [Medline: 19275998]
4. Roehrs A, da Costa CA, da Rosa Righi R, de Oliveira KSF. Personal Health Records: A Systematic Literature Review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
5. Perlin JB. Health Information Technology Interoperability and Use for Better Care and Evidence. JAMA 2016 Oct 25;316(16):1667-1668. [doi: 10.1001/jama.2016.12337] [Medline: 27669026]
6. Tang PC, Lansky D. The missing link: bridging the patient-provider health information gap. Health Aff (Millwood) 2005;24(5):1290-1295 [FREE Full text] [doi: 10.1377/hlthaff.24.5.1290] [Medline: 16162575]
7. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
8. Steele R, Min K, Lo A. Personal health record architectures: Technology infrastructure implications and dependencies. Acta Anaesthesiol Scand 2012 Mar 05;63(6):1079-1091. [doi: 10.1002/asi.22635]
9. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
10. Baird A, North F, Raghu T. Personal Health Records (PHR) and the future of the physician-patient relationship. : ACM International Conference Proceeding Series; 2011 Presented at: iConference 2011; Feb 8-11, 2011; Seattle, WA, USA p. 281-288. [doi: 10.1145/1940761.1940800]
11. Spil T, Klein R. The personal health future. Health Policy and Technology 2015 Jun;4(2):131-136. [doi: 10.1016/j.hlpt.2015.02.004]
12. Pushpangadan S, Seckman C. Consumer perspective on personal health records: A review of the literature. Online Journal of Nursing Informatics 2015 Feb 04;19(1):A [FREE Full text]
13. My Health Record. Personal Health Record Locker URL: https://myhealthrecord.nhp.gov.in/ [accessed 2018-03-25] [WebCite Cache ID 6yBdYdYLI]
14. Research2Guidance. 2016. mHealth App Developer Economics 2016 URL: https://research2guidance.com/r2g/r2g-mHealth-App-Developer-Economics-2016.pdf [accessed 2018-03-25] [WebCite Cache ID 6yBdtUZ4Y]
15. Mandl KD, Kohane IS. A 21st-Century Health IT System - Creating a Real-World Information Economy. N Engl J Med 2017 May 18;376(20):1905-1907. [doi: 10.1056/NEJMp1700235] [Medline: 28514623]
16. PwC. India: PricewaterhouseCooper Private Ltd; 2017. How mHealth can revolutionise the Indian healthcare industry URL: https://www.pwc.in/assets/pdfs/publications/2017/how-mhealth-can-revolutionise-the-indian-healthcare-industry.pdf [accessed 2018-06-10] [WebCite Cache ID 6zLhayWH9]
17. Mandl KD, Mandel JC, Kohane IS. Driving Innovation in Health Systems through an Apps-Based Information Economy. Cell Syst 2015 Jul;1(1):8-13 [FREE Full text] [doi: 10.1016/j.cels.2015.05.001] [Medline: 26339683]
18. Russel J. India's Practo raises $55M to expand its healthcare platform to more emerging markets. 2017 Jan 17. Tech Crunch URL: https://techcrunch.com/2017/01/16/practo-raises-55-million-series-d-led-by-tencent/ [WebCite Cache ID 6zLjMYq6S]
19. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2016 Oct 3:81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]

20. Bedoya A. Comments of Alvaro Bedoya, Center on Privacy & Technology at Georgetown Law, on.: Georgetown Law Aug 15, 2014 URL: https://www.ftc.gov/system/files/documents/public_comments/2014/08/00024-92434.pdf [WebCite Cache ID 6zLkCHpVW]
21. 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]
22. Halamka JD, Tripathi M. The HITECH Era in Retrospect. N Engl J Med 2017 Sep 07;377(10):907-909. [doi: 10.1056/NEJMp1709851] [Medline: 28877012]
23. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
24. Ballantyne A, Schaefer GO. Consent and the ethical duty to participate in health data research. J Med Ethics 2018 Jun;44(6):392-396. [doi: 10.1136/medethics-2017-104550] [Medline: 29358219]
25. White J, Briggs J, Mandel J. NIH and ONC Launch the Sync for Science (S4S) Pilot Enabling Individual Health Data Access and Donation. 2016 URL: https://www.healthit.gov/buzz-blog/health-innovation/nih-and-onc-launch-the-sync-for-science-pilot/ [accessed 2018-03-25] [WebCite Cache ID 6yBfHcOrC]
26. Mohanan M, Hay K, Mor N. Quality Of Health Care In India: Challenges, Priorities, And The Road Ahead. Health Aff (Millwood) 2016 Dec 01;35(10):1753-1758. [doi: 10.1377/hlthaff.2016.0676] [Medline: 27702945]
27. Basu S, Andrews J, Kishore S, Panjabi R, Stuckler D. Comparative performance of private and public healthcare systems in low- and middle-income countries: a systematic review. PLoS Med 2012;9(6):e1001244 [FREE Full text] [doi: 10.1371/journal.pmed.1001244] [Medline: 22723748]
28. Anand S, Fan V. The Health Workforce in India: Human Resources for Health Observer Series No. 2016 URL: http://www.who.int/hrh/resources/16058health_workforce_India.pdf [accessed 2018-03-25] [WebCite Cache ID 6yBfugwBk]
29. Central Bureau of Health Intelligence, Ministry of Health and Family Welfare. National Health Profile 2017 URL: http://www.indiaenvironmentportal.org.in/files/file/NHP_2017-1.pdf [accessed 2018-05-20] [WebCite Cache ID 6zZBDDuAP]
30. Balsari S, Simon G, Nair R, Saunik S, Phadke M. Task shifting in health care: the risks of integrated medicine in India. Lancet Glob Health 2017 Dec;5(10):e963-e964 [FREE Full text] [doi: 10.1016/S2214-109X(17)30322-4] [Medline: 28911757]
31. Srivastava SK. Adoption of Electronic Health Records: A Roadmap for India. Healthc Inform Res 2016 Oct;22(4):261-269 [FREE Full text] [doi: 10.4258/hir.2016.22.4.261] [Medline: 27895957]
32. Sharma M, Aggarwal H. HER Adoption in India: Potential and the Challenges. Indian Journal of Science and Technology Sep 2016;9(34):1-7.
33. HiMMS Analytics. Stage 6 Hospitals URL: http://www.himssanalyticsasia.org/emram/stage6hospitals.asp [accessed 2018-05-12] [WebCite Cache ID 6zMsQhnc3]
34. Powell AC, Ludhar JK, Ostrovsky Y. Electronic health record use in an affluent region in India: Findings from a survey of Chandigarh hospitals. Int J Med Inform 2017 Dec;103:78-82. [doi: 10.1016/j.ijmedinf.2017.04.011] [Medline: 28551005]
35. Patel V, Parikh R, Nandraj S, Balasubramaniam P, Narayan K, Paul VK, et al. Assuring health coverage for all in India. Lancet 2015 Dec 12;386(10011):2422-2435. [doi: 10.1016/S0140-6736(15)00955-1] [Medline: 26700532]
36. Growth Analysis. Östersund, Sweden: Swedish Agency for Growth Policy Analysis; 2013 Apr. India's Healthcare System: Overview and Quality Improvements URL: http://www.tillvaxtanalys.se/download/18.5d9caa4d14d0347533bcf42e/1430909773826/direct_response_2013_04.pdf [WebCite Cache ID 6yC3wjnu7]
37. Vital Wave Consulting. 2009 May. Health Information Systems in Developing Countries URL: http://www.minsa.gob.pe/ogei/conferenciaops/Recursos/43.pdf [WebCite Cache ID 6yC4EncMb]
38. eMaharashtra. Healthcare & Academics Management & Information System URL: https://bit.ly/2zsRg0e [WebCite Cache ID 6yC5Nc1uu]
39. Natarajan S, Panchanathan V, Bansal PK. Health Management Information System Implemented in Government Hospitals of Tamil Nadu. International Journal of Scientific and Engineering Research 2013 Oct;4(10):282-288 [FREE Full text]
40. Amrita Technologies. A sample of Clients using Amrita HIS URL: http://www.amritatech.com/customers.html [accessed 2018-05-12] [WebCite Cache ID 6zN0OCYCg]
41. Gujarat Medical Services Corporation Limited. E-Aushadhi URL: https://gmscl.gujarat.gov.in/e-aushadhi-eng.htm [accessed 2018-05-12] [WebCite Cache ID 6zMuyNck4]
42. E-Aushadhi. Rajasthan Medical Services Corporation Ltd. e-Aushadhi URL: http://e-aushadhi.rajasthan.gov.in/DWH/startup/loginAction [accessed 2018-05-12] [WebCite Cache ID 6zN0kfvJT]
43. Ministry of Health and Family Welfare, Government of India. 2013 Apr. Recommendations On Electronic Medical Records Standards In India URL: https://mohfw.gov.in/sites/default/files/2453910883998892005 1EHR%20Standards-v5%20Apr%202013.pdf [accessed 2018-03-25] [WebCite Cache ID 6yC4mI90b]
44. Rustagi N, Singh R. Electronic medical record: Time to migrate? Perspect Clin Res 2012 Oct;3(4):143-145 [FREE Full text] [doi: 10.4103/2229-3485.103596] [Medline: 23293762]
45. Dandona L, Sivan YS, Jyothi MN, Bhaskar VSU, Dandona R. The lack of public health research output from India. BMC Public Health 2004 Nov 25;4:55 [FREE Full text] [doi: 10.1186/1471-2458-4-55] [Medline: 15563377]

46. Shah J, Pawaskar A, Kumar S, Kshirsagar N. Outcomes research resources in India: current status, need and way forward. Springerplus 2013;2:518 [FREE Full text] [doi: 10.1186/2193-1801-2-518] [Medline: 24171151]
47. Jain D. LiveMint. 2018 Jan 25. Budget 2018: India's health sector needs more funds and better management URL: https://www.livemint.com/Politics/drnszDrkbt418WpuQEHfZI/Budget-2018-Indias-health-sector-needs-more-funds-and-bett.html [accessed 2018-03-26] [WebCite Cache ID 6yC61SwYi]
48. SNOMED International India. URL: https://www.snomed.org/member/india [accessed 2018-03-26] [WebCite Cache ID 6yC6c3a2b]
49. Adler-Milstein J, Jha A. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]
50. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 05;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
51. Ministry of Health & Family Welfare, Government of India. 2017 Oct. National Health Accounts Estimates for India 2014-2015 URL: https://mohfw.gov.in/sites/default/files/National%20Health%20Accounts%20Estimates%20Report%202014-15.pdf [accessed 2018-03-25] [WebCite Cache ID 6yC7EfdLp]
52. Mandl KD, Kohane IS. Escaping the EHR trap--the future of health IT. N Engl J Med 2012 Jun 14;366(24):2240-2242. [doi: 10.1056/NEJMp1203102] [Medline: 22693995]
53. Ministry of Health and Family Welfare, Government of India. 2016 Dec. Electronic Health Record (EHR) Standards for India URL: https://www.nhp.gov.in/NHPfiles/EHR-Standards-2016-MoHFW.pdf [accessed 2018-03-25] [WebCite Cache ID 6yC7bionK]
54. Ministry of Health and Family Welfare.: Government of India National Health Policy, 2017 URL: https://www.nhp.gov.in/NHPfiles/national_health_policy_2017.pdf [accessed 2018-03-26] [WebCite Cache ID 6yC821ZLM]
55. Kandhari R. The Ken. 2018 Jan 08. India picks a builder for its health data hub URL: https://the-ken.com/story/ihip-health-data-hub/ [accessed 2018-03-26] [WebCite Cache ID 6yC8sExDv]
56. Mandl KD, Kohane IS. Time for a Patient-Driven Health Information Economy? N Engl J Med 2016 Jan 21;374(3):205-208 [FREE Full text] [doi: 10.1056/NEJMp1512142] [Medline: 26789868]
57. Iansiti M, Lakhani K. Harvard Business Review. 2017. The Truth About Blockchain URL: https://hbr.org/2017/01/the-truth-about-blockchain [accessed 2018-03-26] [WebCite Cache ID 6yCGq7zQP]
58. Peterson K, Deeduvanu R, Kanjamala P, Boles K. healthIT. 2016. A Blockchain-Based Approach to Health Information Exchange Networks URL: https://www.healthit.gov/sites/default/files/12-55-blockchain-based-approach-final.pdf [WebCite Cache ID 6zMwHIFuX]
59. Fomkin R, Stenbeck M, Litton J. Federated Databases as a Basis for Infrastructure Supporting Epidemiological Research. 2009 Sep 03 Presented at: 20th International Workshop on Database and Expert Systems Application; 2009; Linz, Austria p. 313-317. [doi: 10.1109/DEXA.2009.56]
60. Statista. Share of mobile phone users that use a smartphone in India from 2014 to 2019 URL: https://www.statista.com/statistics/257048/smartphone-user-penetration-in-india/ [accessed 2018-03-26] [WebCite Cache ID 6zMwqrPsF]
61. Bhuyan A. The Wire. 2018 Jan 18. Aadhaar Isn't Just About Privacy. There Are 30 Challenges the Govt Is Facing in Supreme Court URL: https://thewire.in/215017/aadhaar-privacy-government-supreme-court/ [WebCite Cache ID 6yCHbeC5s]
62. Marcoli C. IBM. 2017 Jul 20. Open banking APIs are open for business URL: https://www.ibm.com/blogs/insights-on-business/banking/open-banking-apis-are-open-for-business/ [accessed 2018-03-26] [WebCite Cache ID 6yCCQUAvu]
63. Brodsky L, Oakes L. McKinsey & Company. 2017 Sep. Data sharing and open banking URL: https://www.mckinsey.com/industries/financial-services/our-insights/data-sharing-and-open-banking [accessed 2018-03-26] [WebCite Cache ID 6yCCffgbb]
64. IndiaStack. What is IndiaStack? URL: http://indiastack.org/about/ [accessed 2018-03-30] [WebCite Cache ID 6yJH4PSGO]
65. Unique Identification Authority of India, Government of India. eKYC Dashboard URL: https://uidai.gov.in/aadhaar_dashboard/ekyc_trend.php [accessed 2018-03-26] [WebCite Cache ID 6yCFXpthx]
66. DigiLocker. National eGovernance Division, Ministry of Electronics & Information Technology (MeitY), Government of India URL: https://digilocker.gov.in/ [accessed 2018-03-26] [WebCite Cache ID 6yCFyKp79]
67. OpenMRS. URL: https://wiki.openmrs.org/ [accessed 2018-03-25] [WebCite Cache ID 6yCGGzPE7]
68. SMART. An App Platform for Healthcare URL: https://smarthealthit.org/ [accessed 2018-03-26] [WebCite Cache ID 6yCGZeFa0]
69. Bartlett J. Boston Business Journal. 2017 Jun 08. Athenahealth buys Palo Alto software firm for $63M URL: https://www.bizjournals.com/boston/news/2017/06/08/athenahealth-buys-palo-alto-software-firm-for-63m.html [accessed 2018-03-25] [WebCite Cache ID 6yJEm7avE]
70. Halamka J. Life as a Healthcare CIO. 2017 Dec 17. The Argonaut Project Charter URL: http://geekdoctor.blogspot.com/2014/12/the-argonaut-project-charter.html [WebCite Cache ID 6yCH27cpZ]

71. Ross R. Managing Security Privacy Risk in Healthcare Applications. 2012 Presented at: 5th Annual OCR/NIST HIPAA Security Rule Conference; June 6, 2012; Washington, DC URL: https://csrc.nist.gov/CSRC/media/Presentations/Managing-Security-and-Privacy-Risk-in-Healthcare-A/images-media/day1-1_rross_risk-management-controls.pdf
72. Mandl KD, Kohane IS. No small change for the health information economy. N Engl J Med 2009 Mar 26;360(13):1278-1281. [doi: 10.1056/NEJMp0900411] [Medline: 19321867]
73. Rothstein MA. Is deidentification sufficient to protect health privacy in research? Am J Bioeth 2010 Sep;10(9):3-11 [FREE Full text] [doi: 10.1080/15265161.2010.494215] [Medline: 20818545]
74. Rothstein MA. Is deidentification sufficient to protect health privacy in research? Am J Bioeth 2010 Sep;10(9):3-11 [FREE Full text] [doi: 10.1080/15265161.2010.494215] [Medline: 20818545]
75. de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD. Unique in the Crowd: The privacy bounds of human mobility. Sci Rep 2013;3:1376 [FREE Full text] [doi: 10.1038/srep01376] [Medline: 23524645]
76. Brickell J, Shmatikov V. The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing. 2008 Aug Presented at: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 08; 2008; Las Vegas, Nevada p. 70-78.
77. de Montjoye YA, Radaelli L, Singh VK, Pentland AS. Identity and privacy. Unique in the shopping mall: on the reidentifiability of credit card metadata. Science 2015 Jan 30;347(6221):536-539. [doi: 10.1126/science.1256297] [Medline: 25635097]
78. Young L. IEEE Spectrum. Computer Scientists Find Bias in Algorithms URL: https://spectrum.ieee.org/tech-talk/computing/software/computer-scientists-find-bias-in-algorithms [accessed 2018-03-26] [WebCite Cache ID 6yCIeryis]
79. Simonite T. MIT Technology Review. 2015 Jul 06. Probing the Dark Side of Google's Ad-Targeting System URL: https://www.technologyreview.com/s/539021/probing-the-dark-side-of-googles-ad-targeting-system/ [WebCite Cache ID 6yJF2siVy]
80. Sweeney L. Discrimination in online ad delivery. Commun. ACM 2013 May 01;56(5):44. [doi: 10.1145/2447976.2447990]
81. Kirchner L. The Atlantic. 2015 Jul 06. When Discrimination Is Baked Into Algorithms URL: https://www.theatlantic.com/business/archive/2015/09/discrimination-algorithms-disparate-impact/403969/ [accessed 2018-03-25] [WebCite Cache ID 6yCJXkcfj]
82. Buolamwini J, Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 2018:1-15.
83. Supreme Court of India Civil Original Jurisdiction. Writ Petition (Civil) No 494 of 2012 URL: http://supremecourtofindia.nic.in/supremecourt/2012/35071/35071_2012_Judgement_24-Aug-2017.pdf [accessed 2018-03-26] [WebCite Cache ID 6yDTIjUpp]
84. Ministry of Electronics and Information Technology, Government of India. 2018. White Paper on Data Protection framework for India URL: http://meity.gov.in/white-paper-data-protection-framework-india-public-comments-invited [accessed 2018-03-25] [WebCite Cache ID 6yDTYva1d]
85. Harvard FXB Center for Health and Human Rights. 2018 Feb 08. Harvard FXB Responds to White Paper on Data Protection Framework for India URL: https://fxb.harvard.edu/2018/02/08/harvard-fxb-responds-to-white-paper-on-data-protection-framework-for-india/ [accessed 2018-03-25] [WebCite Cache ID 6yCI3xOKB]
86. Balsari S, Babha J. Jan 31. 2018. Responses to the White Paper of the Committee of Experts on Data Protection Framework for India URL: https://cdn2.sph.harvard.edu/wp-content/uploads/sites/5/2018/02/RHarvard-FXB-Responses-to-WP-Health-Data-FINAL.pdf [accessed 2018-03-26] [WebCite Cache ID 6yDTzD1r9]
87. Ministry of Health & Family Welfare, Government of India. 2017 Nov. Digital Information Security in Healthcare, Act URL: https://mohfw.gov.in/sites/default/files/R_4179_1521627488625_0.pdf [accessed 2018-03-21] [WebCite Cache ID 6yDUMIarO]
88. Press Information Bureau, Government of India. Comments Invited for Creation of National Health Stack. 2018 Jul 06. URL: http://pib.nic.in/PressReleseDetail.aspx?PRID=1538061
89. Government of India, NITI Aayog. National Health Stack: Strategy and Approach. 2018 Jul. URL: http://www.niti.gov.in/writereaddata/files/document_publication/NHS-Strategy-and-Approach-Document-for-consultation.pdf

Abbreviations

API: application programming interface
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
GDPR: General Data Protection Regulation
HIE: Health Information Exchange
HITECH: Information Technology for Economic and Clinical Health
ID: identification
IHIP: Integrated Health Information Platform

IT: information technology
KYC: know your customer
LOINC: Logical Observation Identifiers Names and Codes
MoHFW: The Ministry of Health and Family Welfare
NHPS: National Health Protection Scheme
PHR: personal health record
SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms

Edited by G Eysenbach; submitted 07.04.18; peer-reviewed by R Chandrasekaran, D Walker, P Yao, I Adeleke; comments to author 02.05.18; revised version received 25.05.18; accepted 09.06.18; published 13.07.18.

Please cite as:
Balsari S, Fortenko A, Blaya JA, Gropper A, Jayaram M, Matthan R, Sahasranam R, Shankar M, Sarbadhikari SN, Bierer BE, Mandl KD, Mehendale S, Khanna T
Reimagining Health Data Exchange: An Application Programming Interface–Enabled Roadmap for India
J Med Internet Res 2018;20(7):e10725
URL: http://www.jmir.org/2018/7/e10725/
doi:10.2196/10725
PMID:30006325

Viewpoint

Methodological Shortcomings of Wrist-Worn Heart Rate Monitors Validations

Francesco Sartor[1], PhD; Gabriele Papini[2], MSc; Lieke Gertruda Elisabeth Cox[1], PhD; John Cleland[3], MD, FRCP

[1]Personal Health, Philips Research, Eindhoven, Netherlands

[2]Department of Electrical Engineering, Technical University Eindhoven, Eindhoven, Netherlands

[3]Institute of Health & Well-Being, Robertson Centre for Biostatistics and Clinical Trials, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Francesco Sartor, PhD
Personal Health
Philips Research
High Tech Campus
Eindhoven,
Netherlands
Phone: 31 61 550 9627
Email: francesco.sartor@philips.com

Abstract

Wearable sensor technology could have an important role for clinical research and in delivering health care. Accordingly, such technology should undergo rigorous evaluation prior to market launch, and its performance should be supported by evidence-based marketing claims. Many studies have been published attempting to validate wrist-worn photoplethysmography (PPG)-based heart rate monitoring devices, but their contrasting results question the utility of this technology. The reason why many validations did not provide conclusive evidence of the validity of wrist-worn PPG-based heart rate monitoring devices is mostly methodological. The validation strategy should consider the nature of data provided by both the investigational and reference devices. There should be uniformity in the statistical approach to the analyses employed in these validation studies. The investigators should test the technology in the population of interest and in a setting appropriate for intended use. Device industries and the scientific community require robust standards for the validation of new wearable sensor technology.

(J Med Internet Res 2018;20(7):e10108) doi:10.2196/10108

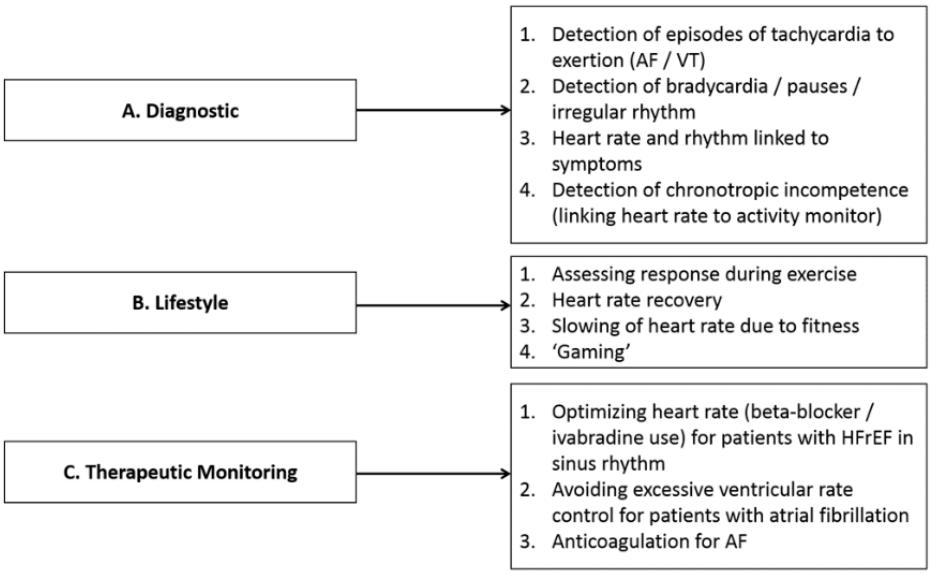
KEYWORDS

sensor technology; accuracy; wearable; telemonitoring

In the past 5 years, there has been a huge proliferation of wrist-worn heart rate monitors, often embedded in smart-bands and smartwatches, which can generate a vast amount of data on lifestyle, physiology, and disease providing exciting opportunities for future health applications. Wearable sensor technology could have an important role for clinical research and in delivering health care [1]. Wearable sensors can be used to encourage healthier living (possible delaying or preventing the onset of disease), screen for incident disease, and provide unobtrusive continuous monitoring for people with chronic illnesses in order to optimize care and detect disease progression and complications. In Figure 1, we show an overview of potential continuous heart rate monitoring applications. New diagnostic applications could become possible thanks to the integration of heart rate and personal information such as age, sex, fitness, activity type, and symptoms. A large number of lifestyle apps and games are emerging thanks to continuous heart rate monitoring, currently most of them related to fitness (eg, Google Fit, Strava) or biofeedback relaxation (eg, Letter Zap, Skip a Beat). It is conceivable that health-promoting apps or games based on heart rate will soon be developed. Wearable heart rate monitors could also enable therapeutic monitoring such as medication titration. Accordingly, such monitors should undergo rigorous evaluation prior to market launch, and their performance should be supported by evidence-based marketing claims [1].

There are several types of validation studies. These studies may be marketing claim validations or medical claim validations for medical grade certification. They are usually done by the manufacturers, sometimes in collaboration with clinical sites, on unreleased products. There may also be benchmarking validation studies, where several commercially available competing products are compared to one another and against a reference. In some cases, there may be even single device validation studies.

Figure 1. Brief overview of potential clinical and nonclinical applications derivable from continuous heart rate monitoring. AF/VT: atrial fibrillation/ventricular tachycardia; HFrEH: heart failure with reduced ejection fraction.



The latter 2 types are generally performed by academic or clinical centers even though industries often engage in such comparisons as well. The only studies which go through a strict quality regulatory framework are medical claim validation studies for medical grade certification (eg, Food and Drug Administration in the United States, medical CE [Conformité Européene] marking in Europe) [2,3]. As a consequence, many nonmedical devices are released on the market without rigorous validation.

In Europe, the choice on how to position a device is the responsibility of the manufacturer, whereas in the United States, this decision can be overruled if the device is perceived to have potential health risks for the user [4]. Because manufacturers can decide whether or not they wish to comply with medical certification regulations, this inevitably leads to heterogeneity in what validations are done. In our view, the lack of stringent regulations for the release of nonmedical heart rate monitoring devices should not justify the lack of standard requirements for validating this technology. The adoption of such technology by health care professionals could be hampered by their liability in case of adverse events when using commercially available nonmedical devices. The authors of this viewpoint agree with Quinn [4], who suggests "a more pragmatic, risk-based approach," which takes a case-by-case look at commercial solutions that may or may not meet the standards required of medical devices. This approach should be applied to promote technology adoption and at the same time safeguard the safety of end-users. Here, we give an overview of clinical applications exploiting wearable heart rate monitors.

In a Research Letter recently published in JAMA [5], the performance of several commercially available, wrist-worn photoplethysmography (PPG)-based heart rate monitors was reported. The authors concluded that PPG-based monitoring was not suitable "when accurate measurement of heart rate is imperative." The authors of that Research Letter acknowledged their report had limitations, including testing only 1 type of activity (treadmill), only in healthy people, and noncontinuous monitoring. Many other studies have been published validating wrist-worn PPG-based heart rate monitoring devices [6-14] but fail to show consensus in favor of or against the accuracy of this sensing technology.

The authors believe that the reason why many validations did not provide conclusive evidence of the validity of wrist-worn PPG-based heart rate monitoring devices is mostly methodological. Studies conducted by teams with a biomedical engineering background are more concerned with addressing problems like signal synchronization and averaging, while research teams with a sports medicine background are more concerned with target groups and exercise protocols. Moreover, clinicians are primarily interested in apps related to telemonitoring, in-hospital or remote. Each approach has its methodological shortcomings. The aim of this viewpoint is to suggest a more consistent and robust approach to validating monitoring technologies.

When validating heart rate monitoring devices, it is sensible to follow a common definition of accuracy. The American National Standards Institute standard for cardiac monitors, heart rate meters, and alarms defines accuracy as a "readout error of no greater than ±10% of the input rate or ±5 bpm, whichever is greater" [15]. Once accurate heart rate is defined, it is also good to agree on what to use as a gold standard. Electrocardiography (ECG) is the accepted gold standard for heart rate monitoring. Nevertheless, ECG, as with PPG, can be severely affected by artifacts [16]. Yet it is generally accepted that PPG-based heart rate monitoring suffers from inherent drawbacks (eg, more difficult peak detection, higher sensitivity to motion artifacts) compared to ECG-based monitoring [16].

The validation strategy should consider the nature of data provided by investigational devices (ID) and reference devices (RD). Heart rate values are always derived from more complex signals (eg, ECG, PPG). Thus, even when the ID and RD have the same output rate (eg, 1 heart rate value per second) and these outputs are well synchronized, the beats compared may not belong to the same time intervals. The method used to extract

information from the raw data (eg, time domain or frequency domain) and the averaging strategy (eg, interbeat intervals or 5-second periods) of the raw data will determine a specific time lag for each heart rate value. Ideally, researchers should have access to the raw data. This is often not possible, and it should be acknowledged as a limitation.

Researchers should realize that their RD (often an ECG device) will not always be accurate. Unless there is a quality check on the validity of the ECG, a second reference device should be used such as a second ECG-based sensor applied in a different manner (eg, patch versus chest strap) and using a different software algorithm for calculating heart rate. When the two RDs fail to agree, no comparison should be made between RD and ID outputs (Figure 2). As mentioned earlier, even the RD (for example ECG patch or ECG strap) in certain circumstances may suffer from inaccuracy due to artifacts (eg, motion artifacts). Based on our own experience in testing hundreds of subjects, we realized that ECG patches perform particularly badly when the skin under the electrodes is stretched or excessively wet. ECG straps perform rather poorly when the skin gets too dry, the strap loosens up, and for certain anatomical shapes (pectus excavatum). These problems must be reported by the researcher.

Figure 2. Correlation between 3 heart rate (HR) monitoring devices and the electrocardiography (ECG) reference. When the 2 chest straps and the wrist-worn photoplethysmography (PPG) heart rate monitors consistently disagree with the reference, their points depart from the 45-degree line in the same way.

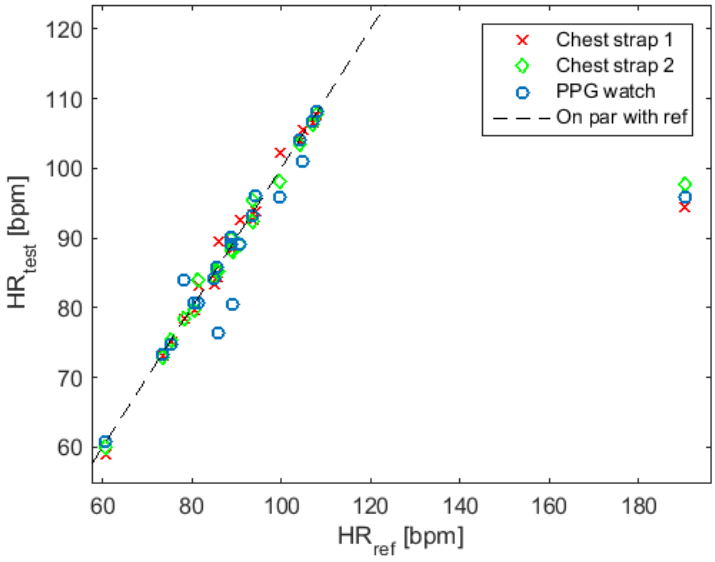
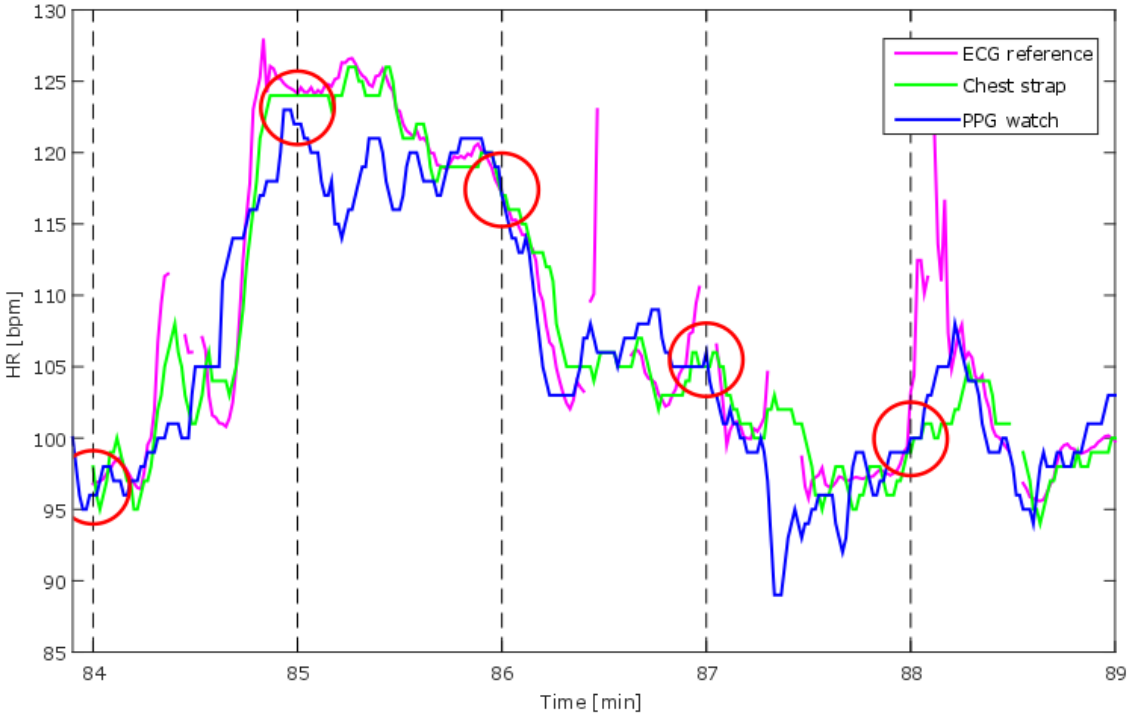


Figure 3. Segment of heart rate (HR) recordings by 3 devices: electrocardiography (ECG) reference, chest strap, and photoplethysmography (PPG) watch. The red circles represent the instants when heart rate from those devices would be collected if these were a value per minute observation. It is evident how these values do not represent the actual second by second or even the average agreement among the 3 devices.



The accuracy of the observation method should be robust (ie, repeatable and reproducible). In some validation studies, heart rate was logged manually after visually consulting the display of both ID and RD [5,7]. This method carries several limitations including human data entry errors and failure to report precisely simultaneous values from multiple devices. This method also limits the observation rate to, for instance, 1 value per minute [5,6]. Taking 1 value per minute is not the same as taking an averaged value over a minute, and both approaches fail to capitalize on the information derived from the rates of change in heart rate and heart variability and assume that participants are in a steady-state condition. Researchers should choose the observation rate (eg, 1 or 5 values per second) and averaging strategy (eg, 5- or 30-second windows) according to the use case foreseen for the heart rate monitor. Yet researchers need to be aware that taking, or averaging, 1 value every minute will hide variability [17]. This is evident in Figure 3, which illustrates that 1 single time point (red circles) is not necessarily representative of the entire minute. Consequently, for the purpose of testing accuracy, even when a mean heart rate value per minute would be sufficient, accuracy should be evaluated at the highest resolution possible.

We also observed a lack of uniformity in the statistical analyses employed in validation studies. Pearson correlations and Student *t* tests are inadequate for testing agreement [18]. This is because the Pearson correlation coefficient is not sensitive to systematic deviations from the 45-degree line, failing to reject agreement when these deviations occur. The Student *t* test is inadequate in rejecting agreement when means are equal but the 2 measures do not correlate with each other, and it can reject agreement when a very small systematic residual error shifts 1 of the means [19]. Moreover, the *t* test assesses difference, which implies that when not rejecting the null hypothesis (ie, means are equal) it does not prove that the 2 means are equivalent. Concordance correlation coefficients should be reported instead [18,19]. Also, limits of agreement analyses should be accompanied by typical error calculations [20]. Equivalence testing should be used when the alternative hypothesis is that the outputs of 2 devices are the same [21]. In equivalence testing, the null hypothesis is that the differences between the means are outside the equivalence limits.

Finally, there are some practical considerations. The investigators should test the technology in the population of interest and in a setting appropriate for intended use. Measurements taken at rest or in the period after exercise cannot be considered to validate measurements done during exercise. Results gathered on healthy individuals with no abnormal heart rhythm are inappropriate for applications aimed at patients with cardiovascular disease where the burden of arrhythmias will be substantially higher. Additionally, due to the effect that the contact of the sensor with the skin and the environmental conditions can have on the PPG signal, information such as sensor placement, strap tightness, skin type, temperature, and possibly light intensity should be reported.

Although many studies have been published to assess the validity and usability of wrist-worn PPG-based heart rate monitoring, their methodological differences and shortcomings hamper research into their clinical utility and their introduction into health care. Such devices could make an important contribution to the future of mobile health and, in our view, should be rigorously evaluated as outlined above. For the reasons discussed in this viewpoint, we advocate standard requirements generally accepted by both the scientific

community and the device industries in order to provide a fair and consistent validation of new wearable sensor technology.

Acknowledgments

The authors would like to thank Dr Helma de Morree for reviewing the first draft of the manuscript.

Conflicts of Interest

FS and LGEC work for Royal Philips Electronics. GP is funded by Stichting voor de Technische Wetenschappen/Instituut voor Innovatie door Wetenschap en Technologie in the context of the Obstructive Sleep Apnea+ project (No 14619). JC has no conflicts of interest.

References

1. Sperlich B, Holmberg H. Wearable, yes, but able...? it is time for evidence-based marketing claims!. Br J Sports Med 2016 Dec 16;51(16):1240 [FREE Full text] [doi: 10.1136/bjsports-2016-097295] [Medline: 27986762]
2. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance).: Official Journal of the European Union; 2017 May 05. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745&from=EN [accessed 2018-06-08]
3. US Code of Federal Regulations, Title 21: Food and Drugs, Part 820—Quality System Regulation, Subpart C—Design Controls. URL: https://www.ecfr.gov/cgi-bin/text-idx?SID=6f0a2b9ffb8a3ac8fe2619b2381bc725&mc=true&node=se21.8.820_130&rgn=div8 [accessed 2018-06-08]
4. Quinn P. The EU commission's risky choice for a non-risk based strategy on assessment of medical devices. Comput Law Security Rev 2017 Jun;33(3):361-370. [doi: 10.1016/j.clsr.2017.03.019]
5. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2016 Oct 12;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
6. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The accuracy of heart rate monitoring by some wrist-worn activity trackers. Ann Intern Med 2017 Apr 18;166(8):610-612. [doi: 10.7326/L16-0353] [Medline: 28395305]
7. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
8. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
9. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
10. Delgado-Gonzalo R, Parak J, Tarniceriu A, Renevey P, Bertschi M, Korhonen I. Evaluation of accuracy and reliability of PulseOn optical heart rate monitoring device. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:430-433. [doi: 10.1109/EMBC.2015.7318391] [Medline: 26736291]
11. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photoplethysmography. Conf Proc IEEE Eng Med Biol Soc 2014;2014:3670-3673. [doi: 10.1109/EMBC.2014.6944419] [Medline: 25570787]
12. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):1 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
13. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
14. Valenti G, Westerterp K. Optical heart rate monitoring module validation study. 2013 Presented at: IEEE International Conference on Consumer Electronics; 2013; Las Vegas p. 195-196. [doi: 10.1109/ICCE.2013.6486856]
15. ANSI/AAMI. Cardiac Monitors, Heart Rate Meters, and Alarms. Arlington: American National Standards Institute, Inc; 2002.
16. Lang M. Beyond Fitbit: a critical appraisal of optical heart rate monitoring wearables and apps, their current limitations and legal implications. Albany Law J Sci Technol 2017;28(1):39-72.
17. Guidance for Industry, Investigating Out-of-Specification (OOS), Test Results for Pharmaceutical Production.: Department of Health and Human Services, Center for Drug Evaluation and Research (CDER); 2006. URL: https://www.fda.gov/downloads/drugs/guidances/ucm070287.pdf [accessed 2018-06-08] [WebCite Cache ID 701iuEKAU]

18. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol 2013 Jun 05;166(1):15-29. [doi: 10.1016/j.ijcard.2012.03.119] [Medline: 22809539]
19. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989 Mar;45(1):255-268. [Medline: 2720055]
20. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000 Jul;30(1):1-15. [Medline: 10907753]
21. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Soc Psychol Personal Sci 2017 May;8(4):355-362 [FREE Full text] [doi: 10.1177/1948550617697177] [Medline: 28736600]

Abbreviations

ECG: electrocardiography
ID: investigational device
PPG: photoplethysmography
RD: reference device

Edited by G Eysenbach; submitted 13.02.18; peer-reviewed by M Lang, P Wark; comments to author 22.03.18; revised version received 16.05.18; accepted 29.05.18; published 02.07.18.

Please cite as:
Sartor F, Papini G, Cox LGE, Cleland J
Methodological Shortcomings of Wrist-Worn Heart Rate Monitors Validations
J Med Internet Res 2018;20(7):e10108
URL: http://www.jmir.org/2018/7/e10108/
doi:10.2196/10108
PMID:29967000

Review

Impact of Internet-Based Interventions on Caregiver Mental Health: Systematic Review and Meta-Analysis

Diana Sherifali[1,2,3*], RN, BSN, PhD; Muhammad Usman Ali[4*], MSc, MD; Jenny Ploeg[5,6,7*], RN, PhD; Maureen Markle-Reid[5,7,8*], RN, MSN, PhD; Ruta Valaitis[5,7,9,10*], RN, PhD; Amy Bartholomew[5*], RN, BSN, MSN; Donna Fitzpatrick-Lewis[4*], MSW; Carrie McAiney[11*], PhD

[1]School of Nursing, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[2]McMaster Evidence Review and Synthesis Team, School of Nursing, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[3]Diabetes Care and Research Program, Hamilton Health Sciences, Hamilton, ON, Canada

[4]Department of Health Research Methods, Evidence and Impact, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[5]Aging, Community and Health Research Unit, School of Nursing, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[6]Department of Health, Aging and Society, Faculty of Social Sciences, McMaster University, Hamilton, ON, Canada

[7]McMaster Institute for Research on Aging, McMaster University, Hamilton, ON, Canada

[8]Person-Centred Interventions for Older Adults with Multimorbidity and their Caregivers, Hamilton, ON, Canada

[9]World Health Organization Collaborating Centre for Primary Care and Health Human Resources, School of Nursing, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[10]Department of Family Medicine, McMaster University, Hamilton, ON, Canada

[11]Program for Interprofessional Practice, Education and Research, Department of Psychiatry and Behavioural Neurosciences, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Diana Sherifali, RN, BSN, PhD
School of Nursing
Faculty of Health Sciences
McMaster University
HSC 3N28F
1280 Main Street West
Hamilton, ON,
Canada
Phone: 1 905 525 9140 ext 21435
Fax: 1 905 524 5199
Email: dsherif@mcmaster.ca

Abstract

Background: The health of informal caregivers of adults with chronic conditions is increasingly vital since caregivers comprise a large proportion of supportive care to family members living in the community. Due to efficiency and reach, internet-based interventions for informal caregivers have the potential to mitigate the negative mental health outcomes associated with caregiving.

Objective: The objective of this systematic review and meta-analysis was to examine the impact of internet-based interventions on caregiver mental health outcomes and the impact of different types of internet-based intervention programs.

Methods: MEDLINE, EMBASE, CINAHL, PsycINFO, Cochrane, and AgeLine databases were searched for randomized controlled trials or controlled clinical trials published from January 1995 to April 2017 that compared internet-based intervention programs with no or minimal internet-based interventions for caregivers of adults with at least 1 chronic condition. The inclusion criteria were studies that included (1) adult informal caregivers (aged 18 years or older) of adults living in the community with a chronic condition; (2) an internet-based intervention program to deliver education, support, or monitoring to informal caregivers; and (3) outcomes of mental health. Title and abstract and full-text screening were completed in duplicate. Data were extracted by a single reviewer and verified by a second reviewer, and risk of bias assessments were completed accordingly. Where possible, data for mental health outcomes were meta-analyzed.

Results: The search yielded 7923 unique citations of which 290 studies were screened at full-text. Of those, 13 studies met the inclusion criteria; 11 were randomized controlled trials, 1 study was a controlled clinical trial, and 1 study comprised both study designs. Beneficial effects of any internet-based intervention program resulted in a mean decrease of 0.48 points (95% CI –0.75 to –0.22) for stress and distress and a mean decrease of 0.40 points (95% CI –0.58 to –0.22) for anxiety among caregivers. For studies that examined internet-based information and education plus professional psychosocial support, the meta-analysis results showed small to medium beneficial effect sizes of the intervention for the mental health outcomes of depression (–0.34; 95% CI –0.63 to –0.05) and anxiety (–0.36; 95% CI –0.66 to –0.07). Some suggestion of a beneficial effect on overall health for the use of information and education plus combined peer and professional support was also shown (1.25; 95% CI 0.24 to 2.25). Overall, many studies were of poor quality and were rated at high risk of bias.

Conclusions: The review found evidence for the benefit of internet-based intervention programs on mental health for caregivers of adults living with a chronic condition, particularly for the outcomes of caregiver depression, stress and distress, and anxiety. The types of interventions that predominated as efficacious included information and education with or without professional psychological support, and, to a lesser extent, with combined peer and psychological support. Further high-quality research is needed to inform the effectiveness of interactive, dynamic, and multicomponent internet-based interventions.

Trial Registration: PROSPERO CRD42017075436; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=75436 (Archived by WebCite at http://www.webcitation.org/709M3tDvn)

(J Med Internet Res 2018;20(7):e10668) doi:10.2196/10668

KEYWORDS

internet; support; education; mental health; caregivers; chronic conditions

Introduction

The number of adults living with chronic conditions is increasing globally [1]. Many adults with chronic conditions rely on family or friend caregivers for support [2]. In Canada, it is reported that more than one-quarter of individuals provided support and care for a friend or family member with a long-term health condition, disability, or age-associated issue in a given year [3]. Caregivers supporting family members living with chronic conditions who need assistance with day-to-day functioning play an essential role for families but also for the health care system, as they provide up to 90% of the medical and supportive care needs for their care recipients [4,5]. While there are many benefits to caregiving for a family member, there are also detrimental emotional and mental health impacts of caregiving that are increasingly being identified and for which practical solutions are urgently needed [3,6,7].

Recognizing the negative health impacts of caregiving has led to studies to examine effective interventions to support these individuals. While a variety of interventions have been evaluated for their impact on improving the health of caregivers, with beneficial effects [8], there is great interest in the use of technology as a means of achieving positive outcomes. Eysenbach [9] suggests that efficiency of health care delivery through internet interventions may lead to a reduction in health care costs. Further, internet and eHealth may be more accessible to caregivers, especially those in remote and rural areas, resulting in increased equity to access health care [9].

There have been 15 recent systematic or other reviews of technology interventions (eg, internet, telephone) to support informal caregivers of adults with chronic conditions in the community [10-24]. Eight reviews focused on internet-based interventions designed specifically for caregivers [12,17,19-24]. All of these reviews provided evidence of improvements in caregivers' health as a result of internet-based programs. A recent rapid evidence review evaluated the impact of internet-based interventions on outcomes for caregivers of persons with chronic conditions living in the community [24]. Internet-based interventions resulted in positive effects on mental health outcomes including decreasing depressive symptoms, stress or distress, and anxiety [24]. Limitations of these studies were that a meta-analysis was not performed to quantify the magnitude of effect across studies and determine clinical relevance; therefore, the impact of internet-based interventions on mental health outcomes of caregivers is still not clear.

The primary objective of this study was to conduct a systematic review and meta-analysis to assess the impact of internet-based interventions of any type compared to no or minimal internet-based interventions on the mental health of informal caregivers of adults with at least 1 chronic condition living in the community. The secondary objective was to examine whether specific types of internet-based interventions had a beneficial effect on caregiver mental health.

Methods

This systematic review and meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [25].

Population

The population of interest included informal caregivers aged 18 years and older who were currently providing caregiving support to adults aged 18 years and older (ie, care recipients) living in the community with at least 1 chronic condition.

Interventions

Studies selected for this systematic review included those that examined any internet-based modality to deliver an intervention, which could include either a single component program or multimodal program to informal caregivers. An internet-based

program was defined as any Web-based series of instructions, options, plans, lessons, modules, or curricula.

Outcomes

The primary outcome of interest for this systematic review was mental health, specifically including depressive symptoms, stress/distress, anxiety, coping, overall mental health, quality of life, and overall health. A second paper on other caregiver outcomes reported in these studies (eg, self-efficacy, self-esteem, burden) is in progress.

Study Design

Selection Criteria

Studies were included according to the following inclusion criteria: (1) study designs were a randomized controlled trial (RCT) or a controlled clinical trial (CCT), (2) studies examined any internet-based intervention program for informal caregivers of older adults having at least 1 chronic condition living in the community, (3) studies were published between January 1, 1995, and April 19, 2017, (4) studies were published in English, (5) studies reported on at least 1 mental health outcome of interest, (6) studies used any measurement tool to examine the mental health outcomes of interest, and (7) studies in which the comparator or control group received none or minimal internet-based intervention (eg, links to a website for information). There were no restrictions on the nature of chronic conditions of care recipients. Exclusion criteria included all other types of study designs (ie, observational studies, case reports), studies that compared different types of program- or module-specific internet-based interventions, grey or unpublished literature, conference abstracts, and letters or editorials. All study protocols without preliminary results for data extraction were also excluded.

Search Strategy

A peer-reviewed search strategy was developed by 2 research librarians at McMaster University. EMBASE, MEDLINE, PsycINFO, CINAHL, Cochrane, and AgeLine databases were searched for studies published between January 1, 1995 and April 19, 2017. Reference lists of systematic reviews were searched for relevant studies not captured by the initial search. Results were deduplicated, and the citations were uploaded to a secure internet-based platform. More detailed information about the search terms is available in Multimedia Appendix 1.

Selection of Studies

Two reviewers independently selected studies for possible inclusion based on title and abstract review. Studies deemed to have met inclusion criteria by either reviewer then underwent full-text review. Any disagreements were discussed between reviewers, and a third party was involved to help reach consensus as necessary.

Data Extraction and Quality Assessment

Full data extraction, including characteristics of included studies, was completed by 1 reviewer and verified by a second reviewer. Risk of bias found in individual studies was assessed by 1 reviewer and verified by a second reviewer. Risk of bias was assessed using the Cochrane risk of bias framework [26], which evaluates the level of bias for sequence generation, allocation concealment, blinding, completeness of outcome assessment, selective reporting, and other biases. The quality of the clinical evidence was critically appraised by 1 reviewer and verified by a second reviewer using the Grading of Recommendations Assessment, Development, and Evaluation system (GRADE), which evaluates the risk for bias, inconsistency, indirectness, and imprecision for each outcome [27]. Disagreements were resolved through consensus between the 2 reviewers.

Data Analysis

A meta-analysis was used to combine the results across studies by outcome using the published data from included studies. To perform the meta-analysis, we used immediate posttreatment data (mean, SD) for continuous outcomes such as depression, stress or distress, anxiety, coping, overall mental health, quality of life, and overall health. We used intention-to-treat outcome data where possible; however, if no intention-to-treat data were reported, we used study completer's outcome data.

The DerSimonian and Laird random effects models with inverse variance method were used to generate the summary measures of effect in the form of standardized mean difference (SMD) [28]. The SMD accounts for similar outcomes measured using different assessment tools (eg, depressive symptoms were assessed using different outcome measures such as Center for Epidemiologic Studies Depression Scale and Beck Depression Inventory) [29]. In this situation, it was necessary to standardize the results of the studies to a uniform scale before they could be combined in a quantitative synthesis. SMDs were calculated using change from baseline data for intervention and control groups for each study with relevant outcome data. For each outcome, data from the corresponding study were used to calculate the mean difference between pretreatment (baseline) and posttreatment (final or end point) values along with its SD for both intervention and control groups. In studies where the SD was not reported, we calculated the SD from the reported standard error (SE) of the mean, 95% confidence intervals (CIs) and *P* values or *z* scores using equations provided in Chapter 7 and Chapter 9 of the Cochrane Handbook for Systematic Reviews of Interventions [30,31]. The SMD is interpreted based on its magnitude according to Cohen *d* recommended thresholds (~0.2=small effect, ~0.5=medium effect, ~0.8=large effect) [32].

The primary meta-analysis was to examine any type of internet-based intervention program by mental health outcome. The secondary meta-analysis was to examine the effects of specific types of internet-based intervention programs on mental health outcomes. Based on our previous work [24], intervention types were categorized accordingly: (1) internet-based information or education only, (2) internet-based information or education plus peer psychosocial support (PPS), (3) internet-based information or education plus professional psychosocial support (PFPS), (4) internet-based information or education plus combined peer and professional psychosocial support, and (5) internet-based intervention with telephone monitoring along with combined peer and professional psychosocial support.

Statistical heterogeneity of combined studies was examined using standard methods. The I^2 statistic was used to quantify the magnitude of statistical heterogeneity between studies, where I^2 of 30% to 60% represents moderate and I^2 of >60% represents substantial heterogeneity [33]. A *P* value of <.10 was used as a guide to indicate where statistically significant heterogeneity may exist, upon which a closer examination of study differences was performed. All analyses were performed using the software packages Review Manager (RevMan version 5.3; The Cochrane Collaboration), STATA version 14 (StataCorp LLC), and GRADEpro Guideline Development Tool.

Results

Study Selection

The search resulted in 7923 unique citations that were screened independently by 2 project staff. At title and abstract screening, we excluded 7633 studies, leaving 290 studies to be screened at full-text. Of those 290 studies, we identified 13 studies (14 papers) that met the inclusion criteria for this systematic review. References lists of the on-topic systematic reviews and included studies were searched but no additional studies were added (Figure 1).

Description of Studies

The purpose, methods, participants, intervention, and risk of bias details of the included studies are shown in Multimedia Appendix 2. From among the 13 included studies, there were 11 studies that were RCTs [34-45], 1 study that was a CCT [46], and 1 study that combined both RCT and CCT designs [47]. Five of the included RCTs were conducted in Europe [34-38], and 5 RCTs were conducted in the United States [41-43,45], of which 1 RCT reported relevant outcomes across 2 papers [39,40]. There was 1 RCT conducted in Canada [44]. The CCT was conducted across the United States, Puerto Rico, and Mexico [46], and the combined CCT and RCT was conducted across 3 European countries [47].

In regard to the type of chronic conditions among care recipients, 9 studies included patients with some form of dementia [34-38,41,45-47]. Cardiovascular health disorders were represented in 3 studies, of which 2 studies included only stroke survivors [42,43] and the other study included a mixed stroke population of stroke-related dementia combined with patients having Alzheimer disease and Parkinson disease [44]. One study was based on non–small cell lung cancer care recipients [39,40]. All included studies were considered small in sample size (≤150 subjects per arm) and had a short length of study follow-up (<6 months). One study included a slightly longer study follow-up time period of 1-year [43]. A majority of studies included informal caregivers aged older than 50 years (range 53.8 to 67.8 years) [35-44,47], except in 1 study that included family caregivers who were also partially working and therefore reported a slightly lower age [45]. Two studies did not provide information on the average age of caregivers [34,46]. More than half of the caregivers were female in all of the included studies (range 56.3% to 100%).

From among the 13 included studies, there were 2 studies (15%) that were categorized as having used an internet-based information or education only intervention [41,45], 3 studies (23%) having used an internet-based information or education plus PPS intervention [34,36,37], 1 study (8%) having used an internet-based information or education plus PFPS intervention [35], 6 studies (46%) having used an internet-based information or education plus combined peer and professional psychosocial support intervention [38-40,42-44,46], and 1 study (8%) having used an internet-based intervention with telephone monitoring along with combined peer and professional psychosocial support [47].

Studies had a comparison group defined as receiving no internet-based intervention that could have included minimal guidance on information resources or website use [34,37,38,41,42,44], usual care with or without additional information [36,39,40,43,45,47], printed information [46], or electronic communications (eg, e-bulletins) [35].

Outcome assessment tools used for relevant mental health outcomes varied across studies and are summarized in Multimedia Appendix 3. Among the 13 included studies, outcomes examined included depression (8/13), stress or distress (6/13), anxiety (2/13), coping (2/13), overall mental health (1/13), quality of life (4/13), and overall health (2/13).

Risk of Bias

The results of the critical appraisal of individual studies for level of bias for sequence generation, allocation concealment, blinding, completeness of outcome assessment, selective reporting, and other biases are shown in Figure 2. Overall, the Cochrane Risk of Bias (RoB) showed mixed quality of study methodology: 2 studies with low RoB [35,37], 3 studies with high RoB [36,40,47], and 8 studies with unclear RoB [34,38,41-46].

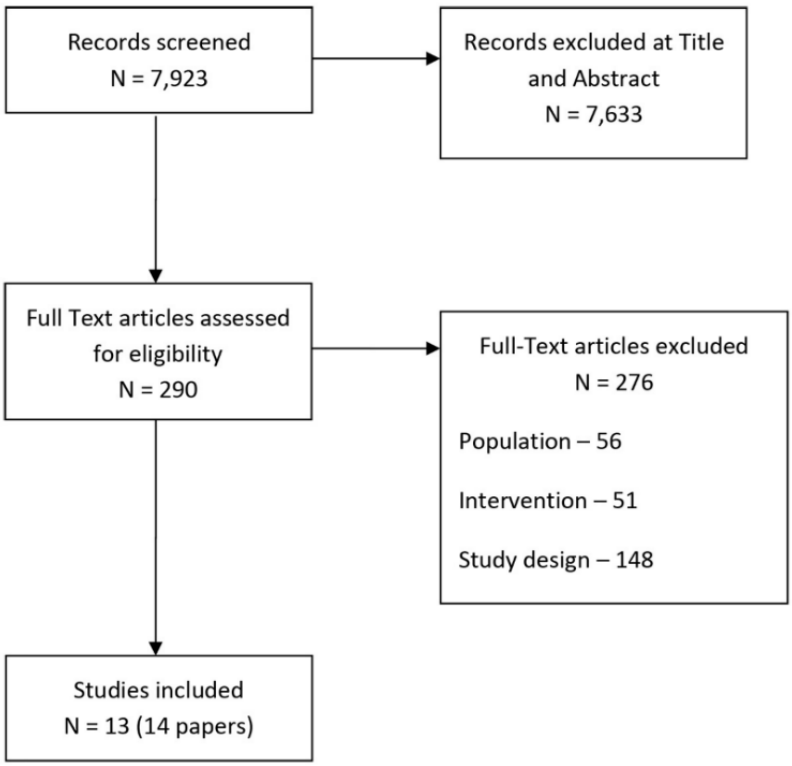
Effectiveness of Internet-Based Interventions

The meta-analysis included an examination of the impact of all internet-based interventions combined as well as an analysis of the impact of each type of internet-based intervention according to mental health outcome. All forest plots are shown in Multimedia Appendix 4.

Any Internet-Based Intervention

A summary of the results of the meta-analysis of any internet-based intervention on mental health outcomes is shown in Table 1. Compared to no or minimal internet-based intervention, any type of internet-based intervention resulted in a beneficial mean decrease of 0.48 points (95% CI –0.75 to –0.22) for stress or distress among caregivers and a beneficial mean decrease of 0.40 points (95% CI –0.58 to –0.22) for anxiety among caregivers. There were no statistically significant differences between groups for the mental health outcomes of depression, coping, overall mental health, quality of life, and overall health. Heterogeneity for the combined effect estimate was observed for the mental health outcomes of depression, stress or distress, quality of life, and overall health (*P*<.10) but not for anxiety and coping. The overall GRADE quality of evidence for each outcome ranged from very low to low.

Figure 1. Study flowchart.



Types of Internet-Based Interventions

Mental health outcomes of interest were examined by the different types of internet-based interventions as shown in Table 2. For information or education only interventions, results showed a beneficial mean decrease of 0.31 points (95% CI –0.50 to –0.11) for depression, a beneficial mean decrease of 0.57 points (95% CI –0.77 to –0.37) for stress or distress, and a beneficial mean decrease of 0.42 points (95% CI –0.65 to –0.19) for anxiety among caregivers, compared to minimal or no internet-based intervention. These results were based on moderate quality of evidence. The remaining mental health outcomes of coping, quality of life, and overall health did not show statistically significant differences between groups. Four of the 6 mental health outcomes of interest included only 1 study. No heterogeneity was detected for the mental health outcomes.

For studies that examined information or education plus PPS, there were no differences between intervention and control groups for any of the mental health outcomes in which there were data including depression, stress or distress, quality of life, and overall health. For studies that included information or education plus PFPS as the intervention, results showed a beneficial mean decrease of 0.34 points (95% CI –0.63 to –0.05) for depression and a beneficial mean decrease of 0.36 points (95% CI –0.66 to –0.07) for anxiety among caregivers, compared to minimal or no internet-based intervention. The quality of evidence for each of these outcomes was moderate.

For studies that examined the intervention of information or education plus combined peer and professional psychological support, results showed a beneficial 1.25-point mean increase for overall health (95% CI 0.24 to 2.25) among caregivers, compared to no or minimal internet-based intervention; however, this result was based on 1 study with an overall sample size of less than 20 caregivers and consequently very low quality of evidence. The remaining outcomes showed no differences between groups. There were no differences between groups for the intervention of information or education with telephone monitoring plus combined peer and professional psychological support for the outcome of quality of life. No other mental health outcomes were reported for this type of intervention. See Multimedia Appendix 5 for the full GRADE assessment details.

Figure 2. Risk of bias for the included studies. U: unclear bias (yellow); L: low risk of bias (green); H: high risk of bias (red).

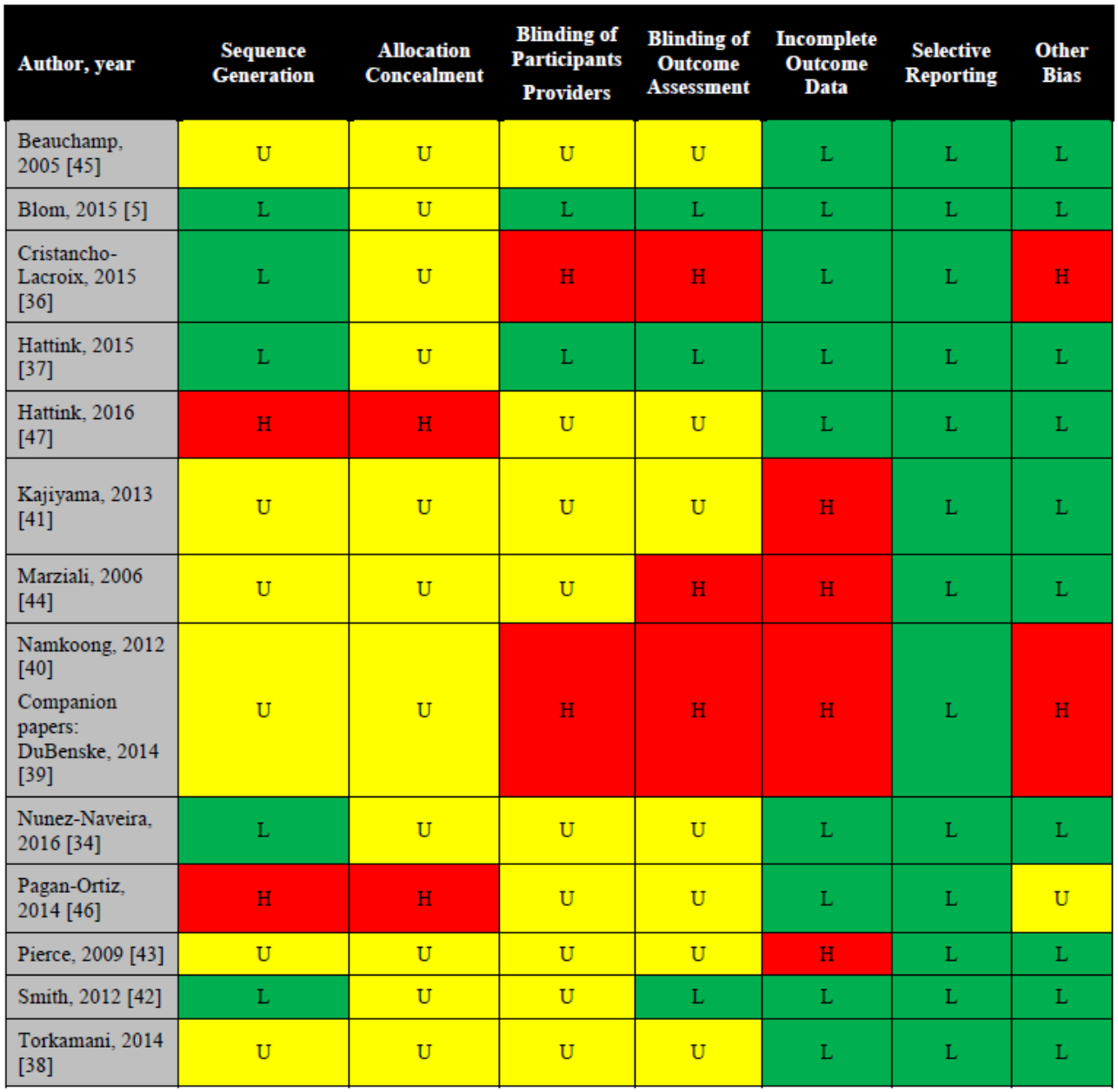


Table 1. Summary of effectiveness of any internet-based interventions.

Mental health outcomes	Number of studies	Intervention/control	Estimate standard mean difference (95% CI)	I^2 (%)	GRADE[a] quality assessment
Depression	8	407/422	–0.19 (–0.43 to 0.05)	59[b]	Very low
Stress or distress	6	288/297	–0.48 (–0.75 to –0.22)	49[b]	Low
Anxiety	2	240/239	–0.40 (–0.58 to –0.22)	0	Low
Coping	2	199/204	–0.01 (–0.20 to 0.19)	0	Very low
Overall mental health	1	45/52	–0.29 (–0.69 to 0.11)	—	Very low
Quality of life	4	102/117	0.01 (–0.49 to 0.51)	68[b]	Very low
Overall health	2	34/34	0.35 (–1.30 to 2.00)	88[b]	Very low

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

[b]Statistically significant heterogeneity (P<.10).

Table 2. Summary of effectiveness of types of internet-based interventions.

Mental health outcomes	Number of studies	Intervention/control	Estimate standard mean difference (95% CI)	I^2 (%)	GRADE[a] quality assessment
Information or education					
Depression	2	196/206	–0.31 (–0.50 to –0.11)	0	Moderate
Stress/distress	2	196/206	–0.57 (–0.77 to –0.37)	0	Moderate
Anxiety	1	150/149	–0.42 (–0.65 to –0.19)	—	Moderate
Coping	1	150/149	0.00 (–0.23 to 0.23)	—	Low
Overall mental health	1	45/52	–0.29 (–0.69 to 0.11)	—	Very low
Quality of life	1	46/57	0.33 (–0.06 to 0.72)	—	Very low
Overall health	1	25/24	–0.44 (–1.01 to 0.13)	—	Very low
Information or education + PPS[b]					
Depression	2	55/55	–0.11 (–0.48 to 0.27)	0	Very low
Stress/distress	2	52/56	–0.46 (–1.41 to 0.50)	83[c]	Very low
Quality of life	1	21/25	–0.36 (–0.95 to 0.22)	—	Very low
Overall health	1	25/24	–0.44 (–1.01 to 0.13)	—	Very low
Information or education + PFPS[d]					
Depression	1	90/90	–0.34 (–0.63 to –0.05)	—	Moderate
Anxiety	1	90/90	–0.36 (–0.66 to –0.07)	—	Moderate
Information or education + combined PPS + PFPS					
Depression	3	66/71	–0.11 (–1.01 to 0.78)	83[c]	Very low
Stress/distress	2	40/35	–0.30 (–1.05 to 0.44)	61[c]	Very low
Coping	1	49/55	–0.03 (–0.41 0.36)	—	Very low
Overall mental health	1	45/52	–0.29 (–0.69 to 0.11)	—	Very low
Quality of life	1	18/20	0.55 (–0.10 to 1.20)	—	Very low
Overall health	1	9/10	1.25 (0.24 to 2.25)	—	Very low
Information or education + telephone monitoring + combined PPS + PFPS					
Quality of life	1	17/15	–0.60 (–1.31 to 0.11)	—	Very low

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

[b]PPS: peer psychosocial support.

[c]Statistically significant heterogeneity (P<.10).

[d]PFPS: professional psychosocial support.

Discussion

Principal Findings

Our systematic review and meta-analysis showed small to moderate beneficial effects of internet-based interventions on caregiver mental health including a reduction in symptoms of depression, stress or distress, and anxiety. The types of internet-based interventions that appeared to have a beneficial effect on mental health included information or education only on decreasing depression, stress or distress, and anxiety and information or education plus PFPS on reducing depression and anxiety. Critical appraisal determined a wide range of the quality of evidence but included a moderate quality of evidence for a modest effect size for a beneficial effect among the 2 specific types of internet-based interventions of information or education only and information or education and PFPS. Additional benefits were shown for the internet-based intervention of information or education plus combined peer and psychological support when it came to overall health among caregivers; however, this was based on a small sample size (<20) and a very low quality of evidence.

Accounting for the type of internet-based intervention revealed additional trends not shown when all types of internet-based interventions were combined. The results showed an approximate 20% increase in the magnitude of effect for stress or distress and an information or education only internet-based intervention among 2 studies, compared to when all 6 studies on stress or distress were combined. Symptoms of depression were improved for an information or education only internet-based intervention as well as for an information or

education plus PFPS internet-based intervention, not shown when all 8 studies on depression were combined.

Based on the detailed critical appraisal and quality assessment of included studies, there are a number of possible reasons that consistent findings across the mental health outcomes were not shown. According to the GRADE scores, the quality of evidence was poor for a number of the outcomes examined, and none of the outcomes was rated as having high-quality evidence. This may reflect, in part, this new and evolving area of focus and the resulting lack of consistency across studies—for example, not all studies examined the mental health outcomes of interest, there was variability in the measurement tools used to assess the different mental health outcomes, the care recipients across studies differed, and too few studies examined the different types of internet-based interventions resulting in small numbers of studies for some outcomes. No differences were noted for multicomponent internet-based interventions on coping and overall mental health since these outcomes were only examined in a few studies. No differences were noted for quality of life perhaps due to small sample sizes and differences in types of interventions. Studies included in the subgroup analyses by type of internet-based intervention were judged to be predominately of low to very low quality of evidence suggesting a number of methodological limitations. Four studies had high risk of bias in the area of incomplete outcome data; 3 studies had high risk of bias for blinding of outcome assessment; and 2 studies had high risk of bias for sequence generation, allocation concealment, and blinding of participants or providers.

There were also many areas where risk of bias could not be assessed due to lack of information in the published papers. For example, risk of bias related to allocation concealment was rated as unclear in 11 of the 13 interventions assessed. Risk of bias related to blinding of participants and providers was rated as unclear in 9 of the 13 interventions examined. Risk of bias related to blinding of outcome assessors was rated as unclear in 7 of the interventions examined. The provision of more detailed information about trial procedures using the Consolidated Standards of Reporting Trials (CONSORT) guidelines for nonpharmacological interventions [48] would enable more accurate assessments of studies for bias and may over time help to elevate the quality of evidence in this area.

We examined the best studies (those with low risk of bias) [35,37,42] to see if there were further insights to be gained. These studies all demonstrated beneficial effects on mental health outcomes: depression [35,42], anxiety [35], stress [37], and quality of life [37]. However, they included quite different types of internet-based interventions. Blom's [35] information or education plus PFPS intervention, targeted at caregivers of persons with dementia, included both a Web-based 8-week course and coaching, monitoring, and evaluation provided by a psychologist. Hattink's [37] information or education plus PPS intervention, also targeted at caregivers of persons with dementia, included a personalized training portal and 2 to 4 months of course materials, interactive exercises, and connection with a Facebook community. Smith's [42] information or education plus combined peer and professional psychosocial support intervention, targeted at spousal caregivers of stroke survivors, involved an 11-week educational program supported by an experienced cardiovascular nurse manager. These interventions had a lengthier intervention period than some of the other studies, 2 of these studies involved professional support, and 1 involved connections with other caregivers. It is possible that these intervention components hold more promise for improving mental health outcomes of caregivers.

Despite significant findings across a range of evidence quality, the intervention mechanism by which improvements in mental health were achieved is still not clear. The interactivity of the information or education only internet-based interventions may have contributed to our significant findings as previously shown by Guay et al [20]. The previously shown important role of human support [20] was variable in our findings, with a beneficial effect shown for the addition of professional psychological support only. It may be that the needs and experiences of the caregivers targeted in these multicomponent interventions are so diverse that the potential impacts of internet-based interventions are not realized. A theoretical basis for internet-based interventions [20,49] has shown to be impactful, and a number of our included studies reported using theory to develop their interventions [36,42,47]. Many interventions included behavior change techniques such as stress management [34,41], problem solving [35], and graded tasks [37], which may have contributed to significant findings. The most efficacious interventions included caregivers and care recipients who were homogeneous, with caregivers characterized as being mostly older female adults and care recipients being those living with some form of dementia [35-38,41,44,45]. Internet-based interventions, when designed with the target populations in mind, may be more likely to demonstrate a beneficial effect on the mental health of caregivers. Internet-based interventions being developed for caregivers should have a strong theoretical basis [50] and incorporate behavior change techniques, particularly those aimed to help manage stress and enhance coping.

Strengths and Limitations

This review summarizes the most relevant trial evidence available to assess the benefits of internet-based interventions on caregiver mental health outcomes. All of the available evidence was published between 2005 and 2017, with more literature published recently (from 2013 to 2017), emphasizing the growing interest in internet technology to support caregivers. However, the review identified that the overall quality of evidence ranged from very low to moderate quality. To our knowledge, this is the first systematic review and meta-analysis examining the impact of internet-based interventions on mental health outcomes of caregivers of adults with chronic conditions living in the community. Although this is an emerging field in the literature, our review set out an a priori selection of rigorous methodological designs, including RCTs and CCTs. This systematic review and meta-analysis was completed with a comprehensive search strategy developed to identify relevant and on-topic literature pertaining to internet-based interventions on informal caregiver mental health outcomes. The review was conducted using methodologically rigorous processes for systematic reviews and meta-analyzed the data using appropriate methods for combining studies that used different outcome assessment tools.

The limitations of the review include the methodological weakness of the studies included, despite being RCTs and CCTs. There was considerable heterogeneity in the interventions across studies. Therefore, we analyzed the impact of the internet-based interventions according to the components of the interventions to understand the impact of these components; however, there were too few studies having used each type of internet-based interventions across all of the mental health outcomes of interest.

Conclusions

This is the first meta-analysis of the impact of internet-based interventions for informal caregivers of adults with chronic conditions on caregiver mental health outcomes. The findings suggest there is an emergence of literature pertaining to internet-based interventions for informal caregivers examining the impact on mental health outcomes. However, future large, high-quality research with clear methodology and consistently reported outcomes of mental health using standardized assessment tools to facilitate meta-analysis and an assessment of clinical relevance are needed to further inform the effectiveness of such interventions, particularly multicomponent internet-based interventions that use peer or professional health care provider support.

Acknowledgments

This work is part of a program of research (Aging, Community, and Health Research Unit) supported by the Canadian Institutes of Health Research Signature Initiative in Community-Based Primary Healthcare (funding reference number: TTF 128261). This work was also supported by the McMaster Evidence Review and Synthesis team.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms.

[PDF File (Adobe PDF File), 50KB - jmir_v20i7e10668_app1.pdf]

Multimedia Appendix 2

Characteristics of included studies.

[PDF File (Adobe PDF File), 104KB - jmir_v20i7e10668_app2.pdf]

Multimedia Appendix 3

Mental health outcomes and measurement tools.

[PDF File (Adobe PDF File), 23KB - jmir_v20i7e10668_app3.pdf]

Multimedia Appendix 4

Meta-analysis and forest plots.

[PDF File (Adobe PDF File), 134KB - jmir_v20i7e10668_app4.pdf]

Multimedia Appendix 5

Detailed Grading of Recommendations, Assessment, Development and Evaluation evidence tables.

[PDF File (Adobe PDF File), 159KB - jmir_v20i7e10668_app5.pdf]

References

1. World Health Organization. 2017. Noncommunicable diseases and their risk factors URL: http://www.who.int/ncds/en/ [accessed 2018-06-13] [WebCite Cache ID 709hQ3v36]
2. Canadian Institute for Health Information. 2011. Seniors and the health care system: what is the impact of multiple chronic conditions? URL: https://secure.cihi.ca/free_products/air-chronic_disease_aib_en.pdf [accessed 2018-06-13] [WebCite Cache ID 709hlVdM4]
3. Sinha M. Statistics Canada. 2013. Spotlight on Canadians: Results from the General Social Survey, Portrait of caregivers, 2012 URL: https://www150.statcan.gc.ca/n1/pub/89-652-x/89-652-x2013001-eng.pdf [accessed 2018-06-13] [WebCite Cache ID 709i6D5LK]

4. Institute of Medicine Committee on the Future Health Care Workforce for Older Americans. Institute of Medicine (US) Committee on the Future Health Care Workforce for Older Americans: Retooling for an Aging America: Building the Health Care Workforce. Washington: National Academies Press; 2008.
5. National Alliance for Caregiving and the AARP.: ; 2004. Caregiving in the US URL: http://www.caregiving.org/data/04finalreport.pdf [accessed 2018-06-13] [WebCite Cache ID 709iXpjiL]
6. Roth DL, Fredman L, Haley WE. Informal caregiving and its impact on health: a reappraisal from population-based studies. Gerontologist 2015 Apr;55(2):309-319. [doi: 10.1093/geront/gnu177] [Medline: 26035608]
7. Turcotte M. Statistics Canada. 2013. Family caregiving:what are the consequences? Insights on Canadian society URL: https://www150.statcan.gc.ca/n1/pub/75-006-x/2013001/article/11858-eng.pdf [accessed 2018-06-13] [WebCite Cache ID 709ijOmd0]
8. Corry M, While A, Neenan K, Smith V. A systematic review of systematic reviews on interventions for caregivers of people with chronic conditions. J Adv Nurs 2015 Apr;71(4):718-734 [FREE Full text] [doi: 10.1111/jan.12523] [Medline: 25223528]
9. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
10. Jackson D, Roberts G, Wu ML, Ford R, Doyle C. A systematic review of the effect of telephone, internet or combined support for carers of people living with Alzheimer's, vascular or mixed dementia in the community. Arch Gerontol Geriatr 2016;66:218-236. [doi: 10.1016/j.archger.2016.06.013] [Medline: 27372903]
11. Lee E. Do technology-based support groups reduce care burden among dementia caregivers? A review. J Evid Inf Soc Work 2015;12(5):474-487. [doi: 10.1080/15433714.2014.930362] [Medline: 25794367]
12. Boots LMM, de Vugt ME, van Knippenberg RJM, Kempen GIJM, Verhey FRJ. 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]
13. Godwin KM, Mills WL, Anderson JA, Kunik ME. Technology-driven interventions for caregivers of persons with dementia: a systematic review. Am J Alzheimers Dis Other Demen 2013 May;28(3):216-222. [doi: 10.1177/1533317513481091] [Medline: 23528881]
14. Llanque SM, Enriquez M. Interventions for Hispanic caregivers of patients with dementia: a review of the literature. Am J Alzheimers Dis Other Demen 2012 Feb;27(1):23-32. [doi: 10.1177/1533317512439794] [Medline: 22467412]
15. Powell J, Chiu T, Eysenbach G. A systematic review of networked technologies supporting carers of people with dementia. J Telemed Telecare 2008;14(3):154-156. [doi: 10.1258/jtt.2008.003018] [Medline: 18430288]
16. McKechnie V, Barker C, Stott J. Effectiveness of computer-mediated interventions for informal carers of people with dementia: a systematic review. Int Psychogeriatr 2014 Oct;26(10):1619-1637 [FREE Full text] [doi: 10.1017/S1041610214001045] [Medline: 24989249]
17. Kaltenbaugh DJ, Klem ML, Hu L, Turi E, Haines AJ, Hagerty LJ. Using Web-based interventions to support caregivers of patients with cancer: a systematic review. Oncol Nurs Forum 2015 Mar;42(2):156-164. [doi: 10.1188/15.ONF.156-164] [Medline: 25806882]
18. Cheng HY, Chair SY, Chau JP. The effectiveness of psychosocial interventions for stroke family caregivers and stroke survivors: a systematic review and meta-analysis. Patient Educ Couns 2014 Apr;95(1):30-44. [doi: 10.1016/j.pec.2014.01.005] [Medline: 24485756]
19. 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;22(e1):e194-e209. [doi: 10.1136/amiajnl-2014-002817] [Medline: 25125686]
20. Guay C, Auger C, Demers L, Mortenson WB, Miller WC, Gélinas-Bronsard D, et al. Components and outcomes of internet-based interventions for caregivers of older adults: systematic review. J Med Internet Res 2017 Sep 19;19(9):e313 [FREE Full text] [doi: 10.2196/jmir.7896] [Medline: 28928109]
21. Waller A, Dilworth S, Mansfield E, Sanson-Fisher R. Computer and telephone delivered interventions to support caregivers of people with dementia: a systematic review of research output and quality. BMC Geriatr 2017 Nov 16;17(1):265 [FREE Full text] [doi: 10.1186/s12877-017-0654-6] [Medline: 29145806]
22. Egan KJ, Pinto-Bruno C, Bighelli I, Berg-Weger M, van Straten A, Albanese E, et al. Online training and support programs designed to improve mental health and reduce burden among caregivers of people with dementia: a systematic review. J Am Med Dir Assoc 2018 Mar;19(3):200-206. [doi: 10.1016/j.jamda.2017.10.023] [Medline: 29306605]
23. Tang WP, Chan CW, So WK, Leung DY. Web-based interventions for caregivers of cancer patients: a review of literatures. Asia Pac J Oncol Nurs 2014;1(1):9-15 [FREE Full text] [doi: 10.4103/2347-5625.135811] [Medline: 27981077]
24. 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]
25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

26. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
27. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]
28. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [Medline: 3802833]
29. Callahan LF, Kaplan MR, Pincus T. The Beck Depression Inventory, Center for Epidemiological Studies Depression Scale (CES-D), and General Well-Being Schedule depression subscale in rheumatoid arthritis. Criterion contamination of responses. Arthritis Care Res 1991 Mar;4(1):3-11. [Medline: 11188584]
30. Higgins J, Deeks J. Selecting studies and collecting data. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. Ottawa: The Cochrane Collaboration; 2011.
31. Deeks J, Higgins J, Altman D. Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. Ottawa: The Cochrane Collaboration; 2011.
32. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. Mahwah: Lawrence Erlbaum Associates; 1988.
33. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. Ottawa: The Cochrane Collaboration; 2011.
34. Núñez-Naveira L, Alonso-Búa B, de Labra C, Gregersen R, Maibom K, Mojs E, et al. UnderstAID, an ICT platform to help informal caregivers of people with dementia: a pilot randomized controlled study. Biomed Res Int 2016:1-13 [FREE Full text] [doi: 10.1155/2016/5726465] [Medline: 28116300]
35. Blom MM, Zarit SH, Groot ZRBM, 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]
36. 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]
37. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J, et al. Web-based STAR E-Learning course increases empathy and understanding in dementia caregivers: results from a randomized controlled trial in the Netherlands and the United Kingdom. J Med Internet Res 2015;17(10):e241 [FREE Full text] [doi: 10.2196/jmir.4025] [Medline: 26519106]
38. Torkamani M, McDonald L, Saez AI, Kanios C, Katsanou M, 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]
39. DuBenske LL, Gustafson DH, Namkoong K, Hawkins RP, Atwood AK, Brown RL, et al. CHESS improves cancer caregivers' burden and mood: results of an eHealth RCT. Health Psychol 2014 Oct;33(10):1261-1272 [FREE Full text] [doi: 10.1037/a0034216] [Medline: 24245838]
40. Namkoong K, DuBenske LL, Shaw BR, Gustafson DH, Hawkins RP, Shah DV, et al. Creating a bond between caregivers online: effect on caregivers' coping strategies. J Health Commun 2012;17(2):125-140 [FREE Full text] [doi: 10.1080/10810730.2011.585687] [Medline: 22004055]
41. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian TY, 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]
42. 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]
43. 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]
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. 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]
46. 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]
47. Hattink BJJ, Meiland FJM, Overmars-Marx T, de Boer M, Ebben PWG, van Blanken M, et al. The electronic, personalizable Rosetta system for dementia care: exploring the user-friendliness, usefulness and impact. Disabil Rehabil Assist Technol 2016 Jul;11(1):61-71. [doi: 10.3109/17483107.2014.932022] [Medline: 24989993]

48. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann Intern Med 2017 Jul 04;167(1):40-47. [doi: 10.7326/M17-0046] [Medline: 28630973]
49. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
50. Pingree S, Hawkins R, Baker T, duBenske L, Roberts LJ, Gustafson DH. The value of theory for enhancing and understanding e-health interventions. Am J Prev Med 2010 Jan;38(1):103-109 [FREE Full text] [doi: 10.1016/j.amepre.2009.09.035] [Medline: 20117565]

Abbreviations

CCT: controlled clinical trial
CI: confidence interval
CONSORT: Consolidated Standards of Reporting Trials
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
PFPS: professional psychosocial support
PPS: peer psychosocial support
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB: risk of bias
SE: standard error
SMD: standardized mean difference

Edited by G Eysenbach; submitted 03.04.18; peer-reviewed by E Marziali, V Cristancho-Lacroix, D Gustafson; comments to author 26.04.18; revised version received 24.05.18; accepted 24.05.18; published 03.07.18.

Please cite as:
Sherifali D, Ali MU, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, McAiney C
Impact of Internet-Based Interventions on Caregiver Mental Health: Systematic Review and Meta-Analysis
J Med Internet Res 2018;20(7):e10668
URL: http://www.jmir.org/2018/7/e10668/
doi:10.2196/10668
PMID:29970358

Original Paper

Traditional and Web-Based Technologies to Improve Partner Notification Following Syphilis Diagnosis Among Men Who Have Sex With Men in Lima, Peru: Pilot Randomized Controlled Trial

Jesse L Clark[1], MSc, MD; Eddy R Segura[1,2], MD, MPH; Catherine E Oldenburg[3], MPH, ScD; Hector J Salvatierra[4], MD; Jessica Rios[4]; Amaya Gabriela Perez-Brumer[5], MSc; Pedro Gonzales[4], MD; Bhupendra Sheoran[6], MBA; Jorge Sanchez[7], MD, MPH; Javier R Lama[4], MD, MPH

[1]David Geffen School of Medicine, Division of Infectious Diseases, Department of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[2]Escuela de Medicina, Universidad Peruana de Ciencias Aplicadas, Lima, Peru

[3]Francis I Proctor Foundation, University of California, San Francisco, San Francisco, CA, United States

[4]Asociacion Civil Impacta Salud y Educación, Lima, Peru

[5]Mailman School of Public Health, Department of Sociomedical Sciences, Columbia University, New York, NY, United States

[6]YTH, San Francisco, CA, United States

[7]Centro de Investigaciones Tecnológicas, Biomédicas y Medioambientales, Universidad Nacional Mayor San Marcos, Lima, Peru

Corresponding Author:
Jesse L Clark, MSc, MD
David Geffen School of Medicine
Division of Infectious Diseases, Department of Medicine
University of California, Los Angeles
10833 Leconte Avenue
CHS 37-121
Los Angeles, CA, 90095
United States
Phone: 1 310 825 3543
Fax: 1 310 825 3632
Email: jlclark@mednet.ucla.edu

Abstract

Background: Patient-initiated partner notification (PN) following the diagnosis of a sexually transmitted infection is a critical component of disease control in men who have sex with men (MSM) sexual networks. Both printed and internet-based technologies offer potential tools to enhance traditional partner notification approaches among MSM in resource-limited settings.

Objective: This randomized controlled trial aimed to evaluate the effect of 2 different PN technologies on notification outcomes following syphilis diagnosis among MSM in Peru: a Web-based notification system and patient-delivered partner referral cards.

Methods: During 2012-2014, we screened 1625 MSM from Lima, Peru, for syphilis infection and enrolled 370 MSM with symptomatic primary or secondary syphilis (n=58) or asymptomatic latent syphilis diagnosed by serology (rapid plasma reagin, RPR, and Microhemagglutination assay for Treponema pallidum antibody; n=312). Prior to enrollment, potential participants used a computer-based self-interviewing system to enumerate their recent sexual partnerships and provide details of their 3 most recent partners. Eligible participants were randomly assigned to one of 4 intervention arms: (1) counseling and patient-initiated Web-based PN (n=95), (2) counseling with Web-based partner notification and partner referral cards (n=84), (3) counseling and partner referral cards (n=97), and (4) simple partner notification counseling (control; n=94). Self-reported partner notification was assessed after 14 days among 354 participants who returned for the follow-up assessment.

Results: The median age of enrolled participants was 27 (interquartile range, IQR 23-34) years, with a median of 2 partners (IQR 1-5) reported in the past month. Compared with those who received only counseling (arm 4), MSM provided with access to Web-based partner notification (arms 1 and 2) or printed partner referral cards (arms 2 and 3) were more likely to have notified one or more of their sexual partners (odds ratio, OR, 2.18, 95% CI 1.30-3.66; *P*=.003 and OR 1.68, 95% CI 1.01-2.79; *P*=.045,

respectively). The proportion of partners notified was also higher in both Web-based partner notification (241/421, 57.2%; *P*<.001) and referral card (240/467, 51.4%; *P*=.006) arms than in the control arm (82/232, 35.3%).

Conclusions: Both new Web-based technologies and traditional printed materials support patient-directed notification and improve self-reported outcomes among MSM with syphilis. Additional research is needed to refine the use of these partner notification tools in specific partnership contexts.

Trial Registration: ClinicalTrials.gov NCT01720641; https://clinicaltrials.gov/ct2/show/NCT01720641 (Archived by WebCite at http://www.webcitation.org/70A89rJL4)

(J Med Internet Res 2018;20(7):e232) doi:10.2196/jmir.9821

KEYWORDS

infectious diseases; syphilis; partner notification; men who have sex with men; Latin America

Introduction

Initially developed in the 1930s as a technique for syphilis control in the United States, notification of sexual partners following the diagnosis of a sexually transmitted infection (STI) remains central to efforts aimed at controlling the spread of HIV and other STIs in at-risk populations [1]. By retracing existing transmission pathways, partner notification (PN) offers the possibility to target efforts to the specific sexual networks structuring the spread of disease in a population [2,3]. Partner-based strategies also provide an opportunity to access individuals and sexual networks that remain unaccessed by traditional HIV or STI education and prevention interventions, such as men who have sex with men (MSM) but do not identify as gay or bisexual [4-6]. The recent resurgence in syphilis infection among MSM in the United States and Europe (and the persistently high prevalence of syphilis in the global South) highlights the importance of developing and refining partner-based efforts to identify and treat syphilis transmission networks in high-risk populations [7-9].

In the field of PN, 3 different types of notification are commonly recognized: (1) provider-initiated notification, in which professional counselors and newly diagnosed individuals work together to identify and notify recent partners; (2) patient-directed notification, in which the index case is encouraged to independently notify their partners after a brief counseling intervention; and (3) contract notification, in which the responsibility for notification shifts from patient to provider after a specified time period [10]. While provider notification is generally more effective than patient-directed efforts, resource-limited health systems in low- and middle-income countries (LMIC) often lack the personnel necessary to conduct detailed contact-tracing studies and so depend primarily on patient-based notification [11-14].

To characterize patient-initiated notification systems, Ferreira et al differentiated between simple and enhanced patient referral [10]. In the simple patient referral, PN is encouraged through professional counseling at the time of diagnosis that emphasizes the importance of notifying recent partners. In the enhanced referral, supplementary tools, including educational video- and theory-based counseling interventions, patient-delivered partner treatment, referral cards, or Web-based notification tools, could be used to motivate and support patient-directed notification efforts [15-18]. Although these standardized categories have formed the central paradigm for PN research, additional research is needed to address overlap and intersection between different methods of notification in real-life clinical populations.

Previously, printed partner referral cards have been used in various circumstances and appear to be effective in promoting notification within heterosexual partnerships [19,20]. However, evidence of the effect of referral cards in developing country settings is limited, and their use within MSM partnerships has not, to our knowledge, been evaluated in any previous randomized controlled trials (RCTs) [21-23].

Data the on the use of internet-based PN systems is mixed. Surveys of STI clinic patients in the United States and Australia, as well as MSM internet users in the United States, have found high levels of acceptability for Web-based notification systems [24-26]. In Europe and Australia, provider-managed internet notification has also been found to be useful in supporting PN during routine clinical practice [27-29]. In our recent study of STI-diagnosed MSM from Lima, Peru, the availability of a website that could deliver anonymous notification messages was expected to significantly increase the notification among men who were expected to not inform their partners under existing conditions [30]. In contrast, previous studies from the United States assessing public awareness and the actual use of PN websites among community-based samples of MSM, as well as male and female visitors to STI clinics, found low levels of penetration into these target populations [31,32]. The only RCT data on the effectiveness of Web-based PN among MSM comes from an incomplete study in Seattle (Washington, USA) that found poor uptake and low levels of acceptability among potential recipients of Web-based notification messages [33]. To the best of our knowledge, no other study has conducted head-to-head comparisons of Web-based systems with other PN technologies [34].

To address this gap in knowledge, we assessed the effect of new and traditional PN technologies on self-reported PN outcomes among MSM recently diagnosed with syphilis infection in Lima, Peru. While the Peruvian Ministry of Health maintains detailed guidelines for managing sexual partners of individuals with syphilis (including provider counseling, distribution of referral cards, and home treatment visits for partners of pregnant women), these services are delivered inconsistently in practice [35,36]. The deficiencies of existing partner management systems in Peru can be seen in the findings from a previous study where after 1 year, only 41% of individuals with syphilis

infection had notified any of their partners and 43% had been reinfected after the confirmation of cure [37,38]. We conducted a factorial, RCT to compare the effect of printed partner referral cards and/or access to a Web-based notification system against standard counseling on self-reported PN outcomes among MSM in Lima, Peru, with untreated syphilis infection.

Methods

Study Design

Between November, 2012 and July, 2014, we conducted a four-arm, factorial RCT (NCT01720641) to assess the effect of new PN tools on notification outcomes among MSM in Lima, Peru, with untreated syphilis infection. The 4 arms included were (1) printed referral cards; (2) Web-based PN; (3) both printed referral cards and Web-based notification; and (4) control (standard of care) counseling procedures.

Screening Procedures

Potential participants were recruited from both community and HIV or STI clinic sites by the staff of the Asociacion Civil Impacta Salud y Educacion (Lima, Peru). Both men and transgender women (TW) who reported anal intercourse with at least one male or transgender female partner in the previous 6 months were invited to participate in an STI screening protocol. Participants in the screening study completed a computer-assisted self-administered (CASI) survey that addressed demographic characteristics, history of HIV and STIs, alcohol and drug use, and sexual network characteristics, as well as attitudes, beliefs, and perceived community norms regarding PN for HIV and STIs. The survey also asked for detailed characteristics of participants' 3 most recent sexual partners, including each partner's gender and sexual identity, sexual practices performed with the partner during the last encounter, and the likelihood of notifying the partner in the event of an STI diagnosis. In order to assist with future recall of partner data, participants were asked to identify each of these 3 partners with a nickname or other identifying characteristic (eg, "the guy in the blue shirt from *La Cueva* ").

All participants underwent a physical examination to identify signs of primary or secondary syphilis infection (painless ulcerative lesions on oral, anal, or genital mucosa or macular rash suggestive of secondary syphilis). Following the clinical examination, participants' blood samples were collected to test for syphilis infection using the rapid plasma reagin (RPR) assay (RPRnosticon; Biomérieux) with microhemagglutination assay for Treponema pallidum antibodies (MHA-TP) confirmation (MHA-TP; Organon Teknika) and serial dilution of positive RPR titers. Although all participants were offered free HIV testing, it was not required as a condition of enrollment. The results of laboratory assays were provided within 2 weeks of screening.

Participants with syphilis infection were treated according to the stage of infection, as determined by a study physician following the review of participants' history of syphilis infection, antibiotic treatment, and RPR titer(s). Cases of symptomatic primary, secondary, and early latent syphilis infection were treated with a single intramuscular injection of 4.2 million units penicillin G benzathine. Conversely, cases of asymptomatic late latent infection were treated with 3 weekly injections of 2.4 million IU penicillin G benzathine. However, participants with newly diagnosed HIV infection were referred to local HIV treatment centers for the initiation of free antiretroviral therapy provided by the Peruvian Ministry of Health.

Randomization and Enrollment

We enrolled MSM and TW diagnosed with untreated syphilis (primary, secondary, or latent infection for which the treating physician recommended antibiotic therapy). Participants with symptomatic evidence of primary or secondary infection were enrolled at the initial screening visit, while participants with asymptomatic infection diagnosed by serology were enrolled after receiving the results of their RPR or MHA-TP testing. All participants provided signed informed consent for a study of, "If and how men with an STI inform their recent sexual partners of their diagnosis." Next, we assigned participants in a 1:1:1:1 fashion to one of the 4 study arms according to a predefined 400-subject randomization scheme generated by the first author at the website (www.random.org).

Intervention and Control Procedures

Each randomization envelope contained an assignment to one of the following 4 arms: (1) Web-based PN, (2) referral cards, (3) combined referral cards and Web-based PN, or (4) control. All allocation assignments were concealed in sealed, opaque, sequentially numbered envelopes that were opened in a numerical order by the study counselor at the point of randomization. To ensure visual and physical consistency of sealed randomization envelopes and to maintain concealment prior to allocation, each envelope was filled with a written study arm assignment, 1 PN counseling script, 5 partner referral cards or blank sheets of paper of the same color and consistency as the referral cards, and 1 Web-based PN access card or a blank note card of the same color and consistency. No deviation from the sequential allocation order or wasting of randomization envelopes was reported.

Randomization envelopes for all 4 arms included a standardized script that was read verbatim by the counselor. The counseling script advised the participants about the importance of notifying their recent partner of their STI diagnosis and informed them of the availability of free testing and treatment resources at the study website, as well as at other area health centers. Participants were also reminded that their safety was paramount and so they should not attempt to notify any partner who might react with violence or abuse.

Participants in the control arm did not receive any additional counseling or PN tools. Participants in the referral cards arm were provided with 5 printed cards to be delivered to a maximum of 5 of their recent sexual partners. Each card contained information about the symptoms and sequelae of syphilis infection, as well as the locations and operating hours of local sites offering free or low-cost HIV and STI testing services. Participants in the Web-based PN arm were read a brief script describing the Web-based notification resources available at (URL:http://www.inspot.org; Accessed: 2018-01-11)

created by YTH, a nonprofit organization designed to use technology to promote youth health and wellness. A Spanish language "Peru" section of the inSPOT website was created specifically for this study and not publicized outside of the trial. Website content was not modified after initiation of the trial. In addition to providing anonymous PN messaging services, the Peru section provided information on testing and treatment resources available in major metropolitan areas of the country. Participants in the Web-based referral arms were provided with a note card indicating the website address. Participants in the combination referral cards/Web-based PN arm were provided with both printed referral cards and access to the inSPOT website using the methods described above.

Endpoint Assessment

Participants in all four arms were asked to return to the clinic in 14-21 days for a follow-up evaluation. The prespecified primary endpoint was self-reported PN. At the follow-up visit, participants completed a brief CASI survey to assess how many of their recent partners (from the 30 days before screening) had been notified, as well as whether each of their 3 most recent partners had been notified and received antibiotic treatment. Participants were reminded of the total number of sexual partners that they had reported at the baseline visit and asked to quantify how many of these partners had been notified. To assist with recall of data for the 3 most recent partners, participants were reminded of the nickname or other identifying characteristic they had assigned each partner, as well as partners' gender and sexual identity. Furthermore, survey questions asked whether each partner had been notified and used a 4-point Likert scale to assess the participants' degree of certainty for whether the partner had received the notification message, whether the partner had sought HIV or STI testing, and whether the partner had received any STI treatment. Operational statistics on the use of the inSPOT.org website during the study period were collected by the YTH staff.

Sample Size and Power Calculations

Sample size calculations were based on previous observational studies of Peruvian men and women diagnosed with HIV or STI [12,39]. Assuming a baseline frequency of 56% for notification of any partner, a sample of 100 subjects per arm was projected to have 80% power to detect a 20% increase in the notification of any recent partner(s).

Statistical Analysis

For the preplanned primary analysis, we recategorized the study arms as follows: (1) Web-based PN (arms 1 and 2); (2) referral cards (arms 2 and 3); and (3) control (arm 4). This approach allowed us to maximize the use of limited resources by assessing two different PN tools within a single clinical trial design, although it was not powered to assess for a synergistic interaction between the interventions [40,41]. Descriptive characteristics for each study arm were calculated with medians and interquartile ranges (IQRs) for continuous variables and proportions for categorical variables. The proportion of participants who reported notifying any recent partner was calculated by the study arm. Next, ORs comparing "Web-based PN versus control" and "referral cards versus control" were calculated with a logistic regression model. We calculated the percentage of all partners, all male partners, stable male partners, and casual male partners who were notified by dividing the total number of partners reported per category by the number of partners per category that was notified. Due to the small number of female partners reported, female partners were excluded from the analysis. Then, we compared the percentage of partners notified in the "Web-based notification versus control" and "referral cards versus control" using the Wilcoxon rank-sum test. For analysis of data obtained from the participants' 3 most recent partners, we used a logistic generalized estimating equation model to assess the notification and treatment outcomes for "Web-based notification versus control" and "referral cards versus control." These outcomes specified (1) if the partner was notified, (2) if the participant knows that the partner received the message, (3) if the partner was known to have been tested for HIV and other STIs, and (4) if the partner was known to have received antibiotic treatment (either delivered by the participant or from another source). No interim analyses were conducted. All analyses were intention-to-treat, and all *P* values were two-sided. All statistical analyses were conducted in Stata 14.1 (StataCorp, College Station, TX, USA).

Human Subjects Protections

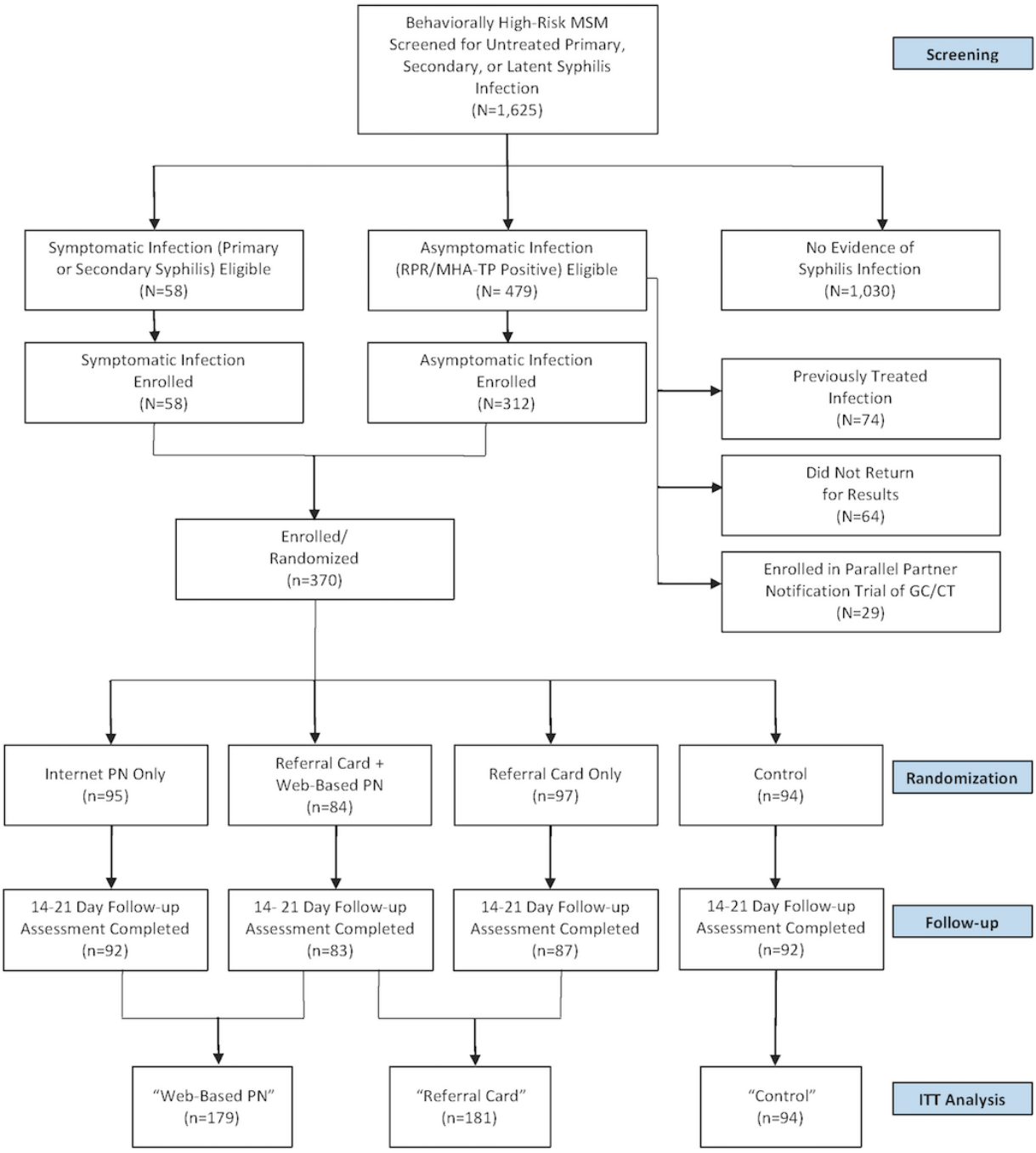
All study procedures were reviewed and approved by the University of California, Los Angeles Office for Human Research Participant Protection (institutional review board #11-003105) and the Asociacion Civil Impacta Comite de Bioetica (Certificate #0052-2012-CE) and were registered with the *Peruvian Instituto Nacional de Salud* before the initiation of any activities. All participants underwent separate informed consent procedures for the screening and RCT protocols and provided written informed consent for each protocol. The clinical trial was registered with clinicaltrials.gov (Protocol Number NCT01720641).

Results

Between November 2012 and June 2014, we screened 1625 individuals, of whom 537 were eligible for enrollment based on results of the physical examination or syphilis serology (Figure 1).

We noted signs and symptoms of primary syphilis in 36 individuals, whereas secondary infection was identified in 22 others. Of 479 MSM and TW with latent infection and positive RPR and MHA-TP assays, 74 were diagnosed with a previously treated infection that did not require additional treatment, 64 did not return for their results, and 29 were enrolled in a concomitant trial of expedited partner therapy for gonorrhea or chlamydia infection [42]. We enrolled 370 MSM or TW with recently diagnosed syphilis infection and randomly assigned each to one of the following four arms: (1) standard counseling or control (arm 1; N=94), (2) referral cards (arm 2; N=97), (3) Web-based PN (arm 3; N=95), or (4) combination Web-based PN and referral cards (arm 4; N=84). According to the prespecified analysis plan, data from arms 2 and 4, as well as arms 3 and 4, were combined into single arms ("referral cards" and "Web-based PN," respectively).

Figure 1. Screening, enrollment, and follow-up (CONSORT) flowchart; Lima, Peru 2012-2014. MSM: men who have sex with men; RPR/MHA-TP: rapid plasma reagin/microhemagglutination Treponema pallidum; GC/CT: gonorrhea/chlamydia; PN: partner notification; ITT: intention-to-treat.



The median age of participants was 30 years in the control group and 27 years in the other two arms (Table 1). The majority of participants had completed secondary school, as well as some university education or technical training, and reported daily or weekly internet use. The most commonly cited sexual identity was "gay or homosexual," with most participants describing their role during intercourse as Moderno (versatile) or Pasivo (receptive). Participants reported a median of 2 male or transgender female sexual partners during the past 30 days. Symptomatic syphilis was more frequently noted among participants in the referral cards (38/181, 21.0%) than the Web-based PN (29/179, 16.2%) and control arms (14/94, 14.8%). Furthermore, the median RPR titer ranged from 1:16 in the Web-based PN arm to 1:32 in the other arms.

At 14-21 Days Follow-up, the proportion of subjects who reported notifying any recent sexual partners of their diagnosis was significantly lower in the Control arm (49/94; 52.1%) than in both the Referral Cards (117/181; 64.6%; Odds Ratio and 95% Confidence Interval: 1.68 [1.01-2.79]) and the Web-based PN (126/179; 70.4%; OR and 95% CI: 2.18 [1.30-3.36]) arms (Table 2).

Table 1. Baseline characteristics by randomization arms (N=370); Lima, Peru, 2012-2014. Arms 1 and 2: Web-Based partner notification (PN; N=179); Arms 2 and 3: Referral Cards (N=181); Arm 4: Control (n=94).

Demographic	Arm 1: Web-based PN only (n=95)	Arm 2: Referral cards + Web-based PN (n=84)	Arm 3: Referral cards only (n=97)	Arm 4: Control (counseling; n=94)
Age (years), median (IQR[a])	27 (23-34)	27 (23-35)	26 (23-30)	30 (24-35)
Education, n (%)				
Primary school only	2 (2.1%)	0 (0%)	0 (0%)	2 (2.2%)
Incomplete secondary school	9 (9.5%)	9 (10.8%)	11 (11.3%)	11 (12.1%)
Complete secondary school	20 (21.0%)	23 (27.4%)	25 (25.8%)	24 (25.5%)
University or vocational training	64 (67.4%)	52 (61.9%)	61 (62.9%)	57 (60.1%)
Sexual identity, n (%)				
Heterosexual	3 (3.2%)	0 (0%)	2 (2.1%)	3 (3.2%)
Bisexual	13 (13.7%)	25 (29.8%)	22 (22.6%)	20 (21.3%)
Homosexual/gay	71 (74.7%)	55 (65.4%)	67 (69.1%)	67 (71.3%)
Trans	2 (2.1%)	1 (1.2%)	0 (0%)	2 (2.1%)
Other	0 (0%)	1 (1.2%)	3 (3.1%)	2 (2.1%)
I don’t know	6 (6.3%)	2 (2.4%)	4 (4.1%)	0 (0%)
Sexual role, n (%)				
Activo (Insertive)	14 (14.7%)	13 (15.5%)	12 (12.4%)	12 (12.8%)
Pasivo (Receptive)	28 (29.5%)	21 (25.0%)	35 (36.0%)	22 (23.4%)
Moderno (Versatile)	47 (49.5%)	49 (58.3%)	43 (44.3%)	57 (60.6%)
Other	4 (4.2%)	1 (1.2%)	3 (3.1%)	2 (2.1%)
I don’t know	2 (6.3%)	0 (0%)	4 (4.1%)	1 (1.1%)
Number of sexual partners (past 30 days), median (IQR)	2 (1-3)	2 (1-3)	3 (1-5)	3 (1-5)
Number of male partners	2 (1-3)	2 (1-4)	3 (1-5)	2 (1-4)
Number of female partners	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)
Frequency of internet use, n (%)				
Daily	69 (72.6%)	64 (76.2%)	68 (70.1%)	69 (73.4%)
Weekly	17 (17.9%)	17 (20.2%)	18 (18.6%)	17 (18.1%)
Monthly	4 (4.2%)	0 (0%)	4 (4.1%)	3 (3.2%)
Less than monthly	3 (3.2%)	1 (1.2%)	2 (2.1%)	2 (2.1%)
Never	2 (2.1%)	2 (2.4%)	5 (5.2%)	3 (3.2%)
RPR[b] titer, median (IQR)	16 (8-64)	16 (8-64)	32 (8-64)	32 (8-64)
Symptomatic infection, n (%)				
Primary syphilis	10 (10.5%)	14 (16.7%)	14 (14.4%)	7 (7.4%)
Secondary syphilis	3 (3.2%)	2 (2.4%)	8 (8.2%)	7 (7.4%)
Latent syphilis	82 (86.3%)	68 (80.9%)	75 (77.4%)	80 (85.2%)

[a]IQR: interquartile range.

[b]RPR: rapid plasma reagin.

Table 2. Partner notification outcomes among men who have sex with men with recently diagnosed syphilis; Lima, Peru; 2012-2014.

	Percentage who notified any recent partners, n (%)	OR[a] (95% CI)	Percentage who notified any recent partners (≥1 recent partner), n (%)	OR (95% CI)
Arm 1: Web-based PN[b] only (n=95)	62/95 (65.2)		62/86 (72.1)	
Arm 2: Referral cards only (n=97)	53/97 (54.6)		53/79 (67.1)	
Arm 3: Referral cards and Web-based PN (n=84)	64/84 (76.2)		64/73 (87.7)	
Arm 4: Control (n=94)	49/94 (52.1)	—	49/79 (62.0)	—
Arms 1+3: All Web-based PN (N=179)	126/179 (70.4)	2.18 (1.30-3.66)	126/159 (79.2)	2.34 (1.29-4.24)
Arms 2+3: All referral cards (N=181)	117/181 (64.6)	1.68 (1.01-2.79)	117/152 (77.0)	2.05 (1.13-3.70)

[a]OR: odds ratio. Arm 4: Control is the reference category for all ORs.

[b]PN: partner notification.

After excluding individuals who denied having any sexual partners in the 30 days before enrollment, the strength of the correlation between the intervention arm and the PN outcomes increased further (referral cards OR 2.05, 95% CI 1.13-3.70; Web-based PN OR 2.34, 95% CI 1.29-4.24).

The proportion of participants' 3 most recent partners who had been notified was significantly higher among those assigned to the referral cards (240/467, 51.4%; 95% CI 44.7%-58.1%; *P*<.001) and Web-based PN (241/421, 57.2%; 95% CI 50.5%-63.9%; *P*<.001) interventions than among those in the control arm (82/232, 35.3%; 95% CI 26.5%-44.1%; Figure 2). This observed difference in self-reported notification outcomes remained significant when limited to all male partners (*P*<.001) and stable male partners (*P*=.01). Although the frequency of notification for casual male partners was significantly higher in both intervention arms, these differences did not attain statistical significance (*P*=.16 and *P*=.54).

The availability of the intervention tools resulted in higher reported frequencies of attempted notification, confirmed notification, and partner STI testing for participants' 3 most recent partners, although none of these comparisons was statistically significant (Figure 3). Furthermore, the frequency of participant-reported partner STI treatment was highest in the Web-based PN arm; however, the proportion of partners known to have received treatment was higher in the control arm than in the referral cards arm.

Figure 2. The proportion of all recent partners notified among men who have sex with men diagnosed with syphilis infection; Lima, Peru 2012-2014. PN: partner notification; *: Calculated using the Rank-Sum Test..

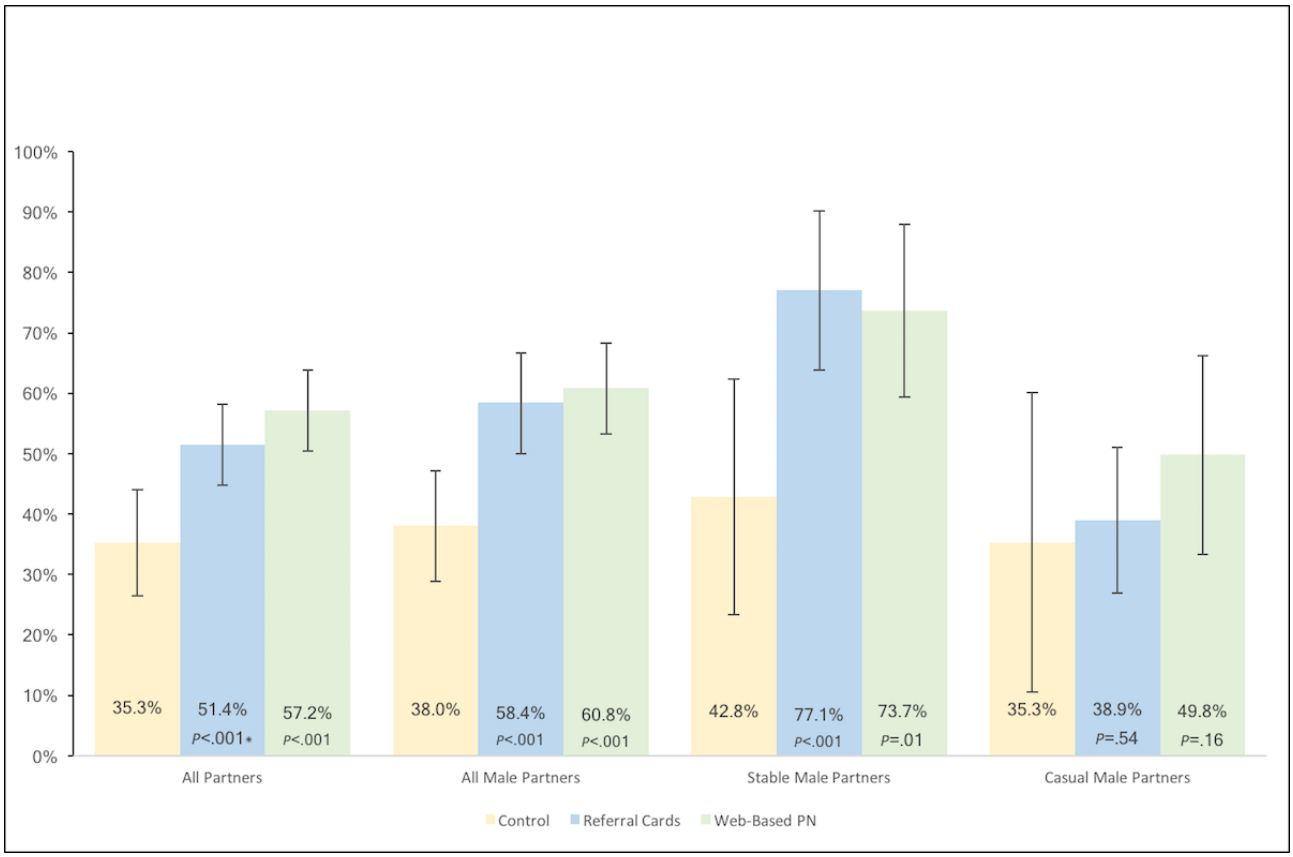
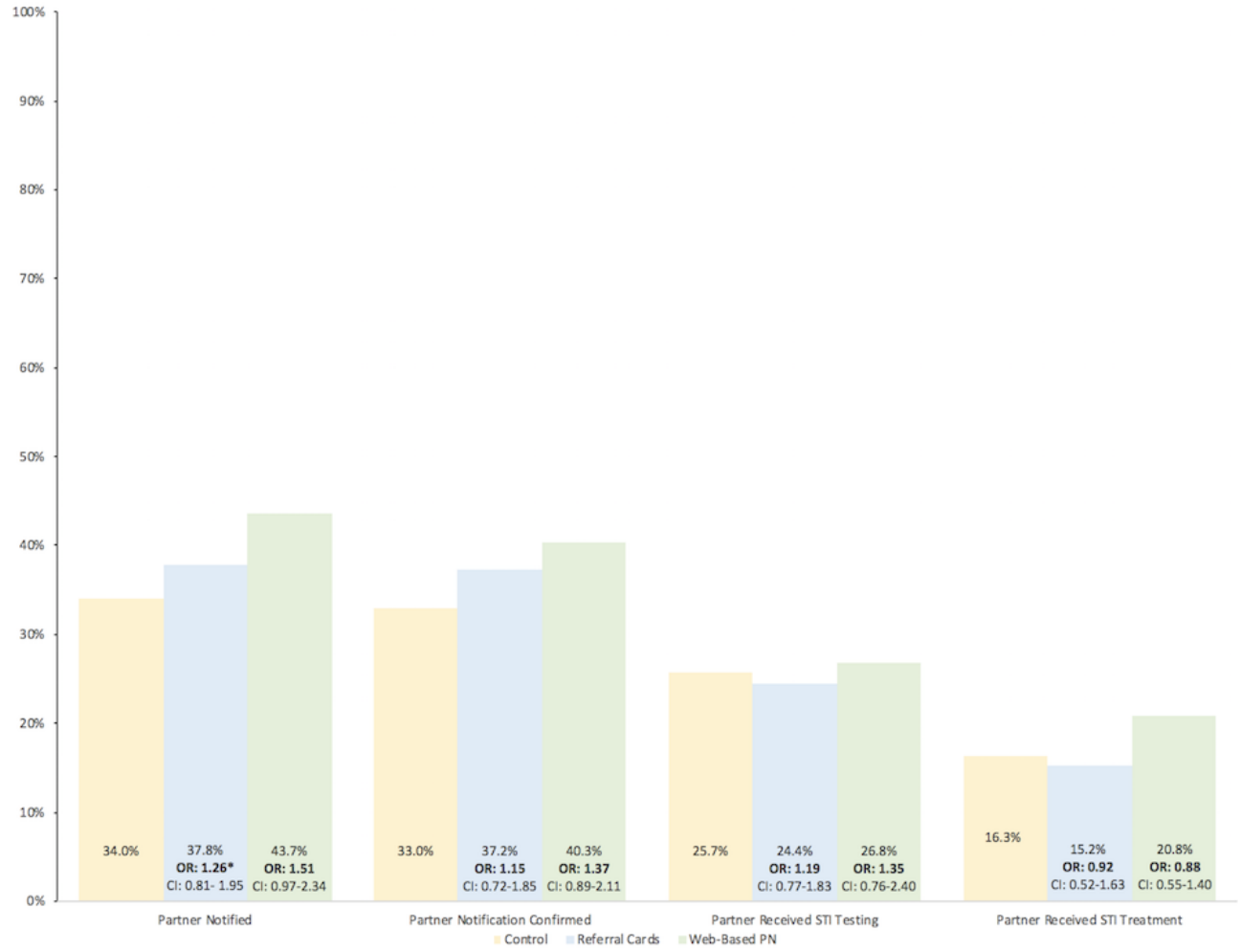


Figure 3. Prevention cascade outcomes of 3 most recent partners of MSM diagnosed with syphilis infection; Lima, Peru 2012-2014. PN: patient notification; STI: sexually transmitted infection. *Generalized Estimating Equation Model.



During the study period, the Peru section of the inSPOT website received 183 unique visits. These visitors sent a total of 47 different e-card messages to 119 recipients (allowing for the possibility of sending the same card to multiple recipients). No episodes of interpersonal violence or partner abuse were reported by study participants in any of the arms.

Discussion

In this study, both new Web-based and traditional printed-media notification tools resulted in significant increases in self-reported PN outcomes compared with simple counseling. Peruvian MSM with newly diagnosed syphilis who were randomly assigned to receive printed referral cards and/or access to a Web-based notification system were more likely to report that they had notified at least one recent sexual partner and to report a larger proportion of recent sex partners. While the greatest impact on notification outcomes was observed among stable partners, smaller increases in notification were also observed with casual partners, particularly among participants randomly assigned to receive access to the Web-based PN system. These findings provide important data to guide global STI control efforts and suggest the potential importance of simple, inexpensive methods to promote patient-directed PN by MSM in resource-limited settings.

To the best of our knowledge, this study is the first to provide RCT evidence of the effectiveness of patient-controlled, Web-based notification systems among MSM. The odds of notifying a recent sexual partner of a potential syphilis exposure were significantly higher among men who received access to the PN website. These results are consistent with findings from our formative research and imply an absolute increase of 30% in the proportion of stable male partners notified and 15% among casual male partners [30]. A synthesis of this data with our formative research findings reveals that a positive effect of Web-based systems on notification outcomes could be attributed to multiple factors, including the availability of an anonymous mechanism to notify partners of a potentially stigmatizing diagnosis, the ability to contact partners for whom only an email address is known, and the opportunity to replicate existing methods of communication among young people already accustomed to communicating through Web-based networks [39,43].

Although extensive research has been conducted on the effect of printed referral cards for PN, this study is one of the first to evaluate their use within MSM partnerships in Latin America.

As with Web-based systems, participants randomly assigned to receive printed partner referral cards exhibited a higher likelihood of notifying at least one partner and a considerably larger proportion of all their recent partners. The observed effect of referral card distribution was highest among stable male partners, with only a minimal effect on the number of casual male partners notified. These findings suggest that notification tools, such as partner referral cards, could be the most valuable in the context of established partnerships, where face-to-face communication is more likely, as opposed to casual sexual contacts where direct, in-person communication is often rare or nonexistent. Taken together, our data support the routine use of these simple, inexpensive tools as resources to enhance patient-driven notification in LMIC settings or other contexts where provider-guided notification programs are unavailable.

In contrast to the substantial impact of both intervention tools on the frequency of notification and the proportion of partners notified, data on more distal outcomes in the notification cascade, such as partner testing and STI treatment, were inconclusive. The assessment of outcomes limited to participants' 3 most recent sexual partners revealed a nonsignificant increase in the participant-reported notification and confirmation of notification, but minor, or even negative, differences in the frequency of partner STI testing and treatment (as reported by participants). As this study was designed to promote and assess only the initial step of notification, the small number of partners included in this section of the analysis and the use of generalized estimating equation modeling to control for multiple observations per participant is likely to have limited our power to detect statistically significant differences. In addition, the dissipation of the intervention effect observed as we progressed through the partner management cascade might reflect the participants' inability to follow up on their notification messages, particularly the ones sent to casual partners through an anonymous, Web-based system. The lack of an observed effect on downstream outcomes indicates the need for additional interventions to support partner management outcomes throughout the partner management cascade, beginning with notification and culminating with a linkage to and retention in HIV or STI care [42]. While the preliminary research on potential responses to hypothetical anonymous notification messages among MSM in Peru indicated that the recipients would be motivated to seek medical attention after receiving an anonymous message, data from the United States have suggested otherwise, and this study too did not collect any partner-confirmed information on postnotification behavior [33,43]. Additional research is needed to evaluate these hypotheses and support both index cases and their partners throughout the HIV or STI prevention and treatment process.

There are several limitations to be considered when interpreting our findings. First, the fact that our data are based on participants' self-report, without independent confirmation by sexual partners, increases the possibility that observed improvements in notification might have been due to social desirability bias in reporting. However, this bias would have likely affected all arms of the study, resulting in a type II error in favor of the null hypothesis and thereby underestimated the true impact of these notification tools. In order to minimize the possibility of desirability bias, we enrolled participants using standardized scripts informing them only that they were invited to a study of, "If and how men with an STI inform their recent sexual partners of their diagnosis," without specifying the different intervention tools being evaluated. In addition, all participants, regardless of the intervention arm, were read the same standardized counseling script advising on the importance of notification and the availability of local testing and treatment resources. Due to the lack of independent partner confirmation, data on the more distal outcomes in the prevention cascade (eg, partner testing and antibiotic treatment) are less reliable and require further investigation with directly confirmed outcomes to be validated. As we did not survey partners on notification outcomes, we were unable assess the cross-contamination between study arms and determine whether individual partners received notification from multiple sources. Similarly, given the relatively small sample size of this study, secondary evaluations of participant- and partner-level factors that might have modified the effect of PN technologies, including the presence or absence of biological symptoms, new diagnosis of HIV coinfection, and differences in the gender and sexual identities of participants and their partners, are beyond the scope of the data presented. Finally, as few TW were enrolled in this study and few cisgender female partners were reported, we did not have sufficient data to draw any conclusions regarding notification by or to MSM and these other groups. Despite these limitations, this study provides important preliminary data to support research into new methods for PN following STI diagnosis in resource-limited settings.

This study provides critical clinical trial evidence to support the effectiveness of both new and traditional notification technologies to support patient-directed PN among MSM with syphilis in Latin America. Regarding the global resurgence in the syphilis incidence within MSM sexual networks, accompanied by endemic levels of transmission among MSM in Latin America, these tools offer simple, inexpensive resources that can dramatically affect the frequency of PN following syphilis diagnosis. While provider-initiated notification by dedicated health professionals is highly effective in promoting PN, testing, and treatment, several resource-limited health systems lack the workforce to implement provider-based notification systems. As a result, strategies to support patient-initiated notification are critical to controlling the disseminated syphilis epidemics that exist among MSM and their male, female, and transgender partners in Latin America. Additional research is needed to explore partner responses to patient-initiated notification messages and develop effective interventions to support testing and treatment outcomes throughout the partner management cascade. Both Web-based notification systems and printed partner referral cards offer simple, effective tools to support the first step in PN, testing, and linkage to care cascade and can fill a unique and essential niche for global HIV and STI prevention efforts.

Acknowledgments

Funding was provided by NIH K23 MH 084611 (PI: Clark), T32 DA 013911 (PI: Flanigan), T32 HD 049339 (PI: Nathanson), R25 MH 087222 (PI: Clark), R25 MH 083620 (PI: Nunn), and R25 TW 009343 (PI: Cohen).

Authors' Contributions

JLC conceptualized and designed the study. JLC, ERS, HJS, JR, PG, BS, JS, and JRL implemented the study procedures. CEO conducted the statistical analysis. JLC, CEO, ERS, and AGPB interpreted the results and prepared the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 696KB - jmir_v20i7e232_app1.pdf]

References

1. Parran T. Shadow on the Land: Syphilis. New York, NY: Reynal and Hitchcock; 1937.
2. Hawkes S, Mabey D, Mayaud P. Partner notification for the control of sexually transmitted infections. BMJ 2003 Sep 20;327(7416):633-634 [FREE Full text] [doi: 10.1136/bmj.327.7416.633] [Medline: 14500409]
3. Hogben M. Partner notification for sexually transmitted diseases. Clin Infect Dis 2007 Apr 01;44 Suppl 3:S160-S174. [doi: 10.1086/511429] [Medline: 17342669]
4. Clark J, Salvatierra J, Segura E, Salazar X, Konda K, Perez-Brumer A, et al. Moderno love: sexual role-based identities and HIV/STI prevention among men who have sex with men in Lima, Peru. AIDS Behav 2013 May;17(4):1313-1328 [FREE Full text] [doi: 10.1007/s10461-012-0210-5] [Medline: 22614747]
5. Rothenberg R. The transformation of partner notification. Clin Infect Dis 2002 Oct 15;35(Suppl 2):S138-S145. [doi: 10.1086/342101] [Medline: 12353200]
6. Wohlfeiler D, Potterat JJ. Using gay men's sexual networks to reduce sexually transmitted disease (STD)/human immunodeficiency virus (HIV) transmission. Sex Transm Dis 2005 Oct;32(10 Suppl):S48-S52. [Medline: 16205293]
7. Hogben M, Paffel J, Broussard D, Wolf W, Kenney K, Rubin S, et al. Syphilis partner notification with men who have sex with men: a review and commentary. Sex Transm Dis 2005 Oct;32(10 Suppl):S43-S47. [Medline: 16205292]
8. Tan WS, Chen M, Ivan M, Stone K, Rane V, Fairley CK, et al. Partner Notification Outcomes for Men Who Have Sex With Men Diagnosed With Syphilis Referred to Partner Notification Officers, Melbourne, Australia. Sex Transm Dis 2016 Dec;43(11):685-689. [doi: 10.1097/OLQ.0000000000000512] [Medline: 27893597]
9. de Lorenzi C, Gayet-Ageron A, Girard-Strohbach M, Toutous-Trellu L. Tracing partners of patients with syphilis infection remains challenging: experience of Geneva Hospital. Int J STD AIDS 2017 Dec;28(11):1090-1097. [doi: 10.1177/0956462416688158] [Medline: 28081685]
10. Ferreira A, Young T, Mathews C, Zunza M, Low N. Strategies for partner notification for sexually transmitted infections, including HIV. Cochrane Database Syst Rev 2013 Oct 03(10):CD002843. [doi: 10.1002/14651858.CD002843.pub2] [Medline: 24092529]
11. Alam N, Chamot E, Vermund SH, Streatfield K, Kristensen S. Partner notification for sexually transmitted infections in developing countries: a systematic review. BMC Public Health 2010 Jan 18;10:19 [FREE Full text] [doi: 10.1186/1471-2458-10-19] [Medline: 20082718]
12. Clark JL, Long CM, Giron JM, Cuadros JA, Caceres CF, Coates TJ, NIMH Collaborative HIV/STD Prevention Trial. Partner notification for sexually transmitted diseases in Peru: knowledge, attitudes, and practices in a high-risk community. Sex Transm Dis 2007 May;34(5):309-313. [doi: 10.1097/01.olq.0000240289.84094.93] [Medline: 17016236]
13. Wang AL, Peng R, Tucker JD, Cohen MS, Chen X. Partner notification uptake for sexually transmitted infections in China: a systematic literature review. Sex Transm Infect 2012 Aug;88(5):386-393 [FREE Full text] [doi: 10.1136/sextrans-2011-050275] [Medline: 22427489]
14. Semple SJ, Pines HA, Strathdee SA, Vera AH, Rangel G, Magis-Rodriguez C, et al. Uptake of a Partner Notification Model for HIV Among Men Who Have Sex With Men and Transgender Women in Tijuana, Mexico. AIDS Behav 2017 Nov 20:1-14. [doi: 10.1007/s10461-017-1984-2] [Medline: 29159592]
15. Bilardi JE, Fairley CK, Hopkins CA, Hocking JS, Temple-Smith MJ, Bowden FJ, et al. Experiences and outcomes of partner notification among men and women recently diagnosed with Chlamydia and their views on innovative resources aimed at improving notification rates. Sex Transm Dis 2010 Apr;37(4):253-258. [doi: 10.1097/OLQ.0b013e3181d012e0] [Medline: 20182407]
16. Rietmeijer C. Using social media for partners services in adolescents. Sex Transm Infect 2013;89(S1):A12.

17. Shiely F, Hayes K, Thomas KK, Kerani RP, Hughes JP, Whittington WLH, et al. Expedited partner therapy: a robust intervention. Sex Transm Dis 2010 Oct;37(10):602-607. [doi: 10.1097/OLQ.0b013e3181e1a296] [Medline: 20601929]
18. Trelle S, Shang A, Nartey L, Cassell JA, Low N. Improved effectiveness of partner notification for patients with sexually transmitted infections: systematic review. BMJ 2007 Feb 17;334(7589):354 [FREE Full text] [doi: 10.1136/bmj.39079.460741.7C] [Medline: 17237298]
19. Kissinger P, Schmidt N, Mohammed H, Leichliter JS, Gift TL, Meadors B, et al. Patient-delivered partner treatment for Trichomonas vaginalis infection: a randomized controlled trial. Sex Transm Dis 2006 Jul;33(7):445-450. [doi: 10.1097/01.olq.0000204511.84485.4c] [Medline: 16531939]
20. Schillinger JA, Kissinger P, Calvet H, Whittington WLH, Ransom RL, Sternberg MR, et al. Patient-delivered partner treatment with azithromycin to prevent repeated Chlamydia trachomatis infection among women: a randomized, controlled trial. Sex Transm Dis 2003 Jan;30(1):49-56. [Medline: 12514443]
21. Díaz OC, Valencia J, Wilson K, García SG, Tinajeros F, Sanchez T. Assessing the effectiveness of a patient-driven partner notification strategy among pregnant women infected with syphilis in Bolivia. Sex Transm Infect 2011 Aug;87(5):415-419. [doi: 10.1136/sti.2010.047985] [Medline: 21460387]
22. Hogben M, Collins D, Hoots B, O Connor K. Partner Services in Sexually Transmitted Disease Prevention Programs: A Review. Sex Transm Dis 2016 Feb;43(2 Suppl 1):S53-S62 [FREE Full text] [doi: 10.1097/OLQ.0000000000000328] [Medline: 26779688]
23. Low N, McCarthy A, Roberts TE, Huengsberg M, Sanford E, Sterne JAC, et al. Partner notification of chlamydia infection in primary care: randomised controlled trial and analysis of resource use. BMJ 2006 Jan 07;332(7532):14-19 [FREE Full text] [doi: 10.1136/bmj.38678.405370.7C] [Medline: 16356945]
24. Hopkins CA, Temple-Smith MJ, Fairley CK, Pavlin NL, Tomnay JE, Parker RM, et al. Telling partners about chlamydia: how acceptable are the new technologies? BMC Infect Dis 2010 Mar 09;10:58 [FREE Full text] [doi: 10.1186/1471-2334-10-58] [Medline: 20211029]
25. Mimiaga MJ, Fair AD, Tetu AM, Novak DS, Vanderwarker R, Bertrand T, et al. Acceptability of an internet-based partner notification system for sexually transmitted infection exposure among men who have sex with men. Am J Public Health 2008 Jun;98(6):1009-1011. [doi: 10.2105/AJPH.2006.098467] [Medline: 17901442]
26. Tomnay JE, Gebert RL, Fairley CK. A survey of partner notification practices among general practitioners and their use of an internet resource for partner notification for Chlamydia trachomatis. Sex Health 2006 Dec;3(4):217-220. [Medline: 17112430]
27. Bilardi JE, Fairley CK, Hopkins CA, Hocking JS, Sze JK, Chen MY. Let Them Know: evaluation of an online partner notification service for chlamydia that offers E-mail and SMS messaging. Sex Transm Dis 2010 Sep;37(9):563-565. [doi: 10.1097/OLQ.0b013e3181d707f1] [Medline: 20351621]
28. Götz HM, van RMS, Vriens P, Op DCE, Hamers M, Heijman T, et al. Initial evaluation of use of an online partner notification tool for STI, called 'suggest a test': a cross sectional pilot study. Sex Transm Infect 2014 May;90(3):195-200. [doi: 10.1136/sextrans-2013-051254] [Medline: 24391062]
29. van RM, Vriens P, Gotz H, Heijman T, Voeten H, Koekenbier R. Acceptance of an onling partner notification tool for STI. called Suggest-a-Test. Sex Transm Infect 2013;89(S1):A335.
30. Clark JL, Segura ER, Perez-Brumer AG, Reisner SL, Peinado J, Salvatierra HJ, et al. Potential impact and acceptability of Internet partner notification for men who have sex with men and transgender women recently diagnosed as having sexually transmitted disease in Lima, Peru. Sex Transm Dis 2014 Jan;41(1):43-45 [FREE Full text] [doi: 10.1097/OLQ.0000000000000068] [Medline: 24326581]
31. Plant A, Rotblatt H, Montoya JA, Rudy ET, Kerndt PR. Evaluation of inSPOTLA.org: an Internet partner notification service. Sex Transm Dis 2012 May;39(5):341-345. [doi: 10.1097/OLQ.0b013e31824e5150] [Medline: 22504593]
32. Rietmeijer CA, Westergaard B, Mickiewicz TA, Richardson D, Ling S, Sapp T, et al. Evaluation of an online partner notification program. Sex Transm Dis 2011 May;38(5):359-364. [doi: 10.1097/OLQ.0b013e31820ef796] [Medline: 21343844]
33. Kerani RP, Fleming M, Golden MR. Acceptability and intention to seek medical care after hypothetical receipt of patient-delivered partner therapy or electronic partner notification postcards among men who have sex with men: the partner's perspective. Sex Transm Dis 2013 Feb;40(2):179-185. [doi: 10.1097/OLQ.0b013e31827adc06] [Medline: 23324981]
34. Hottes TS, Gilbert M. Evaluation of online partner notification services like inSPOT requires starting with the client, not the clinic. Sex Transm Dis 2012 May;39(5):348; author reply 348. [doi: 10.1097/OLQ.0b013e31824c0e77] [Medline: 22504595]
35. García PJ, Williams E, Cárcamo CP, Chiappe M, Holmes KK, Peeling RW, et al. Partner Notification Among Peruvian Pregnant Women With Syphilis. Sex Transm Dis 2015 Aug;42(8):457-462. [doi: 10.1097/OLQ.0000000000000314] [Medline: 26165438]
36. Ministerio de Salud del Peru. Guias Nacionales de Atencion Integral de Salud Sexual y Reproductiva. Guias Nacionales de Atencion Integral de Salud Sexual y Reproductiva 2004:-.
37. Braun HM, Segura ER, Lake JE, Gandhi M, Rios J, Villaran MV, et al. Individual and partnership factors associated with anticipated versus actual partner notification following STI diagnosis among men who have sex with men and/or with

transgender women in Lima, Peru. Sex Transm Infect 2017 Nov 30:-. [doi: 10.1136/sextrans-2017-053292] [Medline: 29191814]
38. Long CM, Klausner JD, Leon S, Jones FR, Giron M, Cuadros J, NIMH Collaborative HIV/STD Prevention Trial Group. Syphilis treatment and HIV infection in a population-based study of persons at high risk for sexually transmitted disease/HIV infection in Lima, Peru. Sex Transm Dis 2006 Mar;33(3):151-155. [doi: 10.1097/01.olq.0000204506.06551.5f] [Medline: 16508525]
39. Clark JL, Perez-Brumer AG, Segura ER, Salvatierra HJ, Sanchez J, Lama JR. Anticipated Notification of Sexual Partners following STD Diagnosis among Men Who Have Sex with Men and Transgender Women in Lima, Peru: A Mixed Methods Analysis. PLoS One 2016;11(9):e0163905 [FREE Full text] [doi: 10.1371/journal.pone.0163905] [Medline: 27685158]
40. McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. JAMA 2003 May 21;289(19):2545-2553. [doi: 10.1001/jama.289.19.2545] [Medline: 12759326]
41. Pocock SJ, Clayton TC, Stone GW. Challenging Issues in Clinical Trial Design: Part 4 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol 2015 Dec 29;66(25):2886-2898 [FREE Full text] [doi: 10.1016/j.jacc.2015.10.051] [Medline: 26718676]
42. Clark JL, Segura ER, Oldenburg CE, Rios J, Montano SM, Perez-Brumer A, et al. Expedited Partner Therapy (EPT) increases the frequency of partner notification among MSM in Lima, Peru: a pilot randomized controlled trial. BMC Med 2017 Dec 04;15(1):94 [FREE Full text] [doi: 10.1186/s12916-017-0858-9] [Medline: 28468648]
43. Clark JL, Perez-Brumer A, Salazar X. "Manejar la Situacion": Partner Notification, Partner Management, and Conceptual Frameworks for HIV/STI Control Among MSM in Peru. AIDS Behav 2015 Dec;19(12):2245-2254 [FREE Full text] [doi: 10.1007/s10461-015-1049-3] [Medline: 25821149]

Abbreviations

CASI: computer-based self-interviewing
GC/CT: gonorrhea/chlamydia
IQR: interquartile range
ITT: intention-to-treat
LMIC: low- and middle-income countries
MHA-TP: microhemagglutination Treponema pallidum
MSM: men who have sex with men
PN: partner notification
RCT: randomized controlled trials
RPR: rapid plasma reagin
STI: sexually transmitted infection
TW: transgender women

Edited by G Eysenbach; submitted 14.01.18; peer-reviewed by M Hogben, J Tucker, F Calvo; comments to author 08.02.18; revised version received 28.02.18; accepted 13.03.18; published 03.07.18.

Please cite as:
Clark JL, Segura ER, Oldenburg CE, Salvatierra HJ, Rios J, Perez-Brumer AG, Gonzales P, Sheoran B, Sanchez J, Lama JR
Traditional and Web-Based Technologies to Improve Partner Notification Following Syphilis Diagnosis Among Men Who Have Sex With Men in Lima, Peru: Pilot Randomized Controlled Trial
J Med Internet Res 2018;20(7):e232
URL: http://www.jmir.org/2018/7/e232/
doi:10.2196/jmir.9821
PMID:29970355

Original Paper

Large-Scale Dissemination of Internet-Based Cognitive Behavioral Therapy for Youth Anxiety: Feasibility and Acceptability Study

Sonja March[1,2], PhD (Clin Psych); Susan H Spence[3,4], BSci, PhD (Clin Psych), MBA; Caroline L Donovan[4], PhD (Clin Psych); Justin A Kenardy[5], PhD

[1]Institute for Resilient Regions, University of Southern Queensland, Springfield, QLD, Australia

[2]School of Psychology and Counselling, University of Southern Queensland, Ipswich, QLD, Australia

[3]Australian Institute for Suicide Research and Prevention, Griffith University, Mt Gravatt, QLD, Australia

[4]School of Applied Psychology, Griffith University, Mt Gravatt, QLD, Australia

[5]School of Psychology, University of Queensland, St Lucia, QLD, Australia

Corresponding Author:
Sonja March, PhD (Clin Psych)
Institute for Resilient Regions
University of Southern Queensland
Sinnathamby Blvd
Springfield, QLD, 4300
Australia
Phone: 61 34704434
Email: sonja.march@usq.edu.au

Abstract

Background: Internet-based cognitive behavioral therapy (iCBT) for child and adolescent anxiety has demonstrated efficacy in randomized controlled trials, but it has not yet been examined when disseminated as a public health intervention. If effective, iCBT programs could be a promising first-step, low-intensity intervention that can be easily accessed by young people.

Objective: The objective of our study was to examine the feasibility and acceptability of a publicly available online, self-help iCBT program (BRAVE Self-Help) through exploration of program adherence, satisfaction, and changes in anxiety.

Methods: This study was an open trial involving the analysis of data collected from 4425 children and adolescents aged 7-17 years who presented with elevated anxiety at registration (baseline) for the iCBT program that was delivered through an open-access portal with no professional support. We assessed the program satisfaction via a satisfaction scale and measured adherence via the number of completed sessions. In addition, anxiety severity was assessed via scores on the Children's Anxiety Scale, 8-item (CAS-8) at four time points: baseline, Session 4, Session 7, and Session 10.

Results: Participants reported moderate satisfaction with the program and 30% completed three or more sessions. Statistically significant reductions in anxiety were evident across all time points for both children and adolescents. For users who completed six or more sessions, there was an average 4-point improvement in CAS-8 scores (Cohen *d*=0.87, children; Cohen *d*=0.81, adolescents), indicating a moderate to large effect size. Among participants who completed nine sessions, 57.7% (94/163) achieved recovery into nonelevated levels of anxiety and 54.6% (89/163) achieved statistically reliable reductions in anxiety.

Conclusions: Participant feedback was positive, and the program was acceptable to most young people. Furthermore, significant and meaningful reductions in anxiety symptoms were achieved by many children and adolescents participating in this completely open-access and self-directed iCBT program. Our results suggest that online self-help CBT may offer a feasible and acceptable first step for service delivery to children and adolescents with anxiety.

(J Med Internet Res 2018;20(7):e234) doi:10.2196/jmir.9211

KEYWORDS

adolescent; child; anxiety disorders; cognitive behavioral therapy; eHealth; public health

Introduction

Half of all lifetime mental health disorders begin before the age of 14 years [1], highlighting the importance of early intervention as a strategy for promoting lifelong mental health. Anxiety is one of the most common childhood mental health conditions, with almost 7% of Australian children and adolescents aged 4-17 years meeting the criteria for an anxiety disorder [2,3]. Although anxiety disorders in youth lead to significant impairment [4], they can be treated effectively using cognitive behavioral therapy (CBT) [5]. Unfortunately, only 56% of young people with mental disorders report having used services in the previous 12 months, with only 2.2% accessing specialist child and adolescent mental health care [2]. The pervasiveness of anxiety and the noted barriers to treatment [6,7] highlight the importance and potential value of evidence-based, population-level early interventions.

The current Australian federal government recommendations encourage primary prevention and early intervention across the life span through easy-to-access first-line responses, particularly for children [8]. They further propose the use of digital and low-intensity mental health services to ensure that all Australians have access to care, before crisis, irrespective of their geographical location [8].

BRAVE Self-Help is an online, open-access, self-help intervention for child and adolescent anxiety that addresses the needs of anxious Australian youth using a digital, low-intensity, and population-level model. The BRAVE Self-Help initiative was initially supported by *beyondblue* and commenced in 2014, offering an evidence-based, open-access, online program free of charge to Australian young people and their parents. The self-help program was adapted from BRAVE-ONLINE, a 10-session, internet-based CBT (iCBT) program implemented with brief therapist support. The evidence base for the therapist-assisted program is strong [9-12], and the program is recognized internationally as the only "probably efficacious" iCBT intervention for childhood anxiety [13]. Furthermore, the program assists young people to develop strategies for identifying and managing anxiety-provoking situations using youth-friendly, engaging, and interactive Web-based sessions. The objective of the current initiative was to examine its feasibility and acceptability when disseminated nationally and offered as an open-access, self-help, early intervention program for young Australians without therapist support.

There has been relatively little research examining the implementation of iCBT programs for child and adolescent anxiety in real-world clinical and community contexts (eg, outside university-based research trials); furthermore, there are no formalized guidelines for a large-scale dissemination [14]. One small feasibility study has very recently been reported by Jolstedt et al [15], where an evidence-based iCBT program for anxiety was implemented in a small sample (N=20) of anxious children in an outpatient clinic in Sweden. The program was delivered with therapist support (20 min/week) and included both child and parent involvement. Overall, the program was acceptable (moderate to high satisfaction) to young people, parents, and clinicians, with participants completing, on average, 6 out of 12 modules and half of the sample reaching at least module 4 (exposure). Furthermore, young people showed significant reductions in anxiety symptoms from pre- to posttreatment (Cohen d=1.22). Thus, the study by Jolstedt et al provided some preliminary support for the dissemination of iCBT interventions in real-world settings, although it was limited to a specific outpatient setting rather than national and open dissemination and participants were required to meet strict inclusion criteria. Thus, further research is required to understand the feasibility of disseminating iCBT programs to large groups of children and adolescents.

Despite substantial challenges involved in determining the outcomes of an open-access, real-world service, we examined the impact of the BRAVE Self-Help intervention through a feasibility and acceptability approach. As the program was developed and intended for youth with anxiety, we were specifically interested in those children and adolescents who reported elevated levels of anxiety at enrollment into the self-help program. Our primary aim was to determine the feasibility and acceptability of BRAVE Self-Help when disseminated nationally through open access. Specifically, we evaluated the level of adherence to and satisfaction with the program as well as the extent to which anxiety symptoms changed over the course of the program for those with elevated symptoms. We hypothesized that children and adolescents who adhered to the program would show significant reductions in their self-reported anxiety from baseline to Sessions 4, 7, and 10. We also expected high satisfaction among participants of the program.

Methods

Participants and Procedure

Participants were 4425 anxious young people aged 7-17 years (1473 children aged 7-12 years; 2952 adolescents aged 13-17 years) with mean (SD) age 12.95 (2.97) years; there were 66.39% (2938/4425) females and 31.77% (1406/4425) males, and 1.84% (81/4425) participants identified as another gender category. In terms of residence, 57.45% (2542/4425) participants resided in major cities, with 23.35% (1033/4425) from Inner Regional Australia, 11.21% (496/4425) from Outer Regional Australia, and 2.55% (113/4425) living in remote or very remote Australia (241/4425 [5.44%] provided data that could not be accurately coded).

All participants registered for the program through a website accessible only to Australian families. Then, participants were directed to the program in several ways through (1) self-referral and internet searching, (2) referral from health or education professionals, (3) links hosted on several Australian mental health information sites (eg, Reach Out and Beacon), and (4) direct links from the *beyondblue* and *youthbeyondblue* websites. Through the *beyondblue* website, the BRAVE Program was listed as a direct referral for young people who completed an anxiety quiz and scored high on anxiety. In order to promote awareness of the program, introduction letters, flyers, and postcards were sent to schools and health and mental health organizations as well as private practitioners nationwide. The program was also presented at relevant conferences held for

school counselors, psychologists, and teachers. Throughout the 2-year recruitment period, 28.67% (1269/4425) of participants were referred by school-based professionals, 13.36% (591/4425) by external health professionals, 10.85% (480/4425) by a parent or family member, 8.48% (375/4425) through *beyondblue*, and 9.94% (440/4425) through internet searching, with the remaining (28.70%, 1270/4425) participants finding the program through other means (eg, word-of-mouth, radio, magazine, and advertisements). With respect to the participants referred from health professionals, 13.96% (80/573) were referred by their general practitioner, 55.67% (319/573) by a psychologist, and 9.60% (55/573) by a social worker, with the remaining participants referred by other health professionals.

To be included in this study, participants were required to have enrolled in the BRAVE Self-Help program between July 1, 2014, and June 30, 2016. We monitored the program progress for participants through to November 17, 2016. All participants (including those registered at the end of the recruitment period) were monitored over a 20-week period, which was sufficient to complete the ten sessions. Participants who were enrolled outside of the 2-year recruitment period were excluded from this study. As this online program is open-access, young people do not need a referral to register and begin the sessions. Besides, there are no set inclusion criteria for enrollment in and access to the program, and thus, users are able to access the program for prevention, early intervention, or treatment purposes. When registering, users are not required to demonstrate symptomatic levels of anxiety, and the program is completely open access and self-sought. In this study, however, participants were included only if they demonstrated elevated anxiety at the baseline above a predetermined criterion (≥84th percentile or *T*-score ≥60 on the Children's Anxiety Scale [CAS-8]; see below). The progression of participants from registration through the study, with reasons for exclusion is presented in Figure 1.

There are two versions of the program, one for children aged 7-12 years and another for teenagers aged 13-17 years; participants selected which version they wish to complete. Informed consent (and parental consent in the case of children aged <16 years) was required prior to beginning the program, and it was obtained during the Web-based registration process. Participation was voluntary, and young people were made aware that they could cease using the program at any time, without consequence. The study protocol approval was obtained from the ethics committees of the University of Queensland (UQ), University of Southern Queensland, and Griffith University. Furthermore, data were stored on secure servers hosted by UQ. Participants were not provided with any reimbursements for participation.

Clinical Intervention

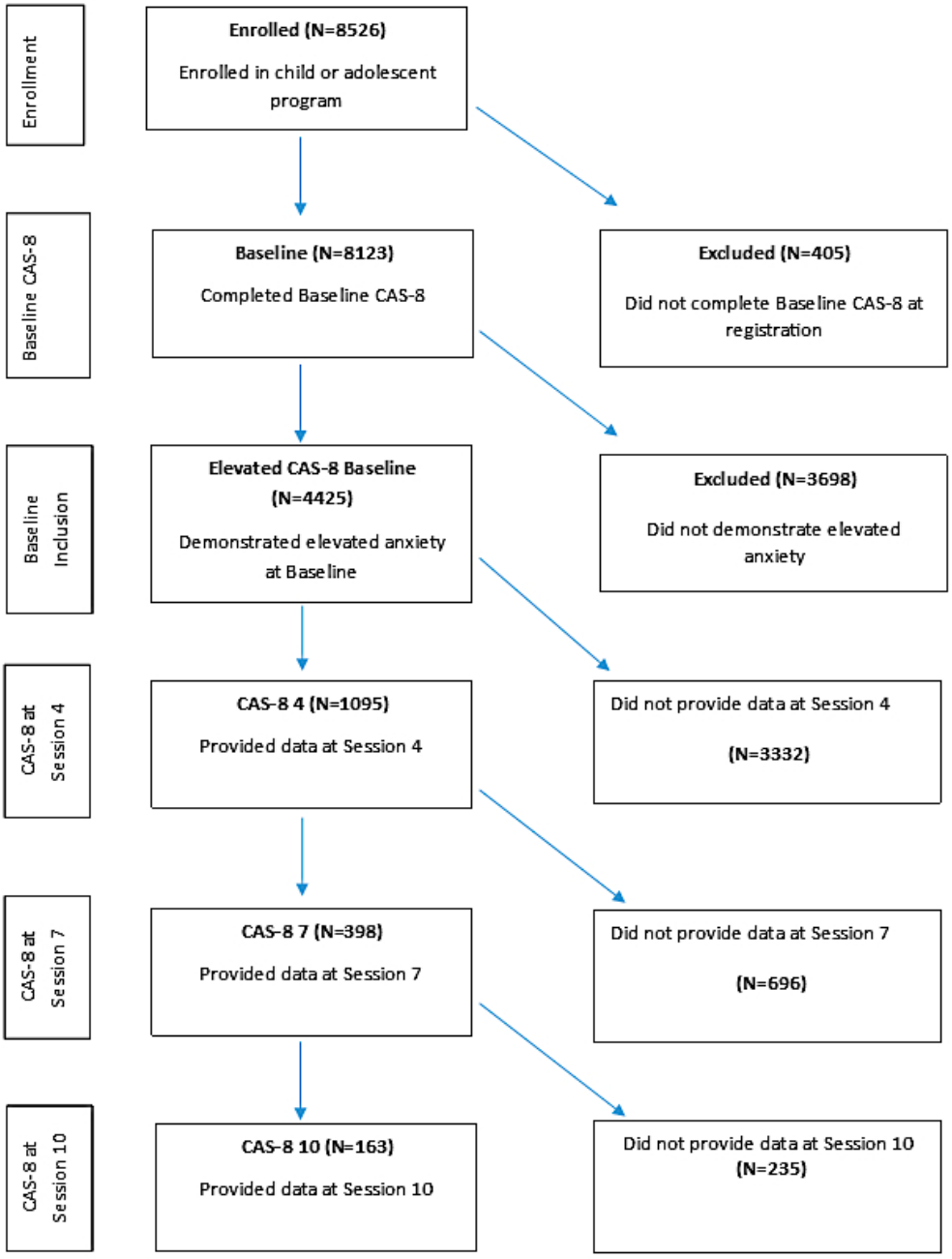
The BRAVE-ONLINE program for youth anxiety when delivered with minimal therapist assistance has been described elsewhere [10-12,16]. For this study, the program content (modules) remained the same as for the therapist-assisted program (including 10 interactive Web-based CBT sessions), although minor adaptations were made to the presentation of material to facilitate the learning and implementation of the complex CBT strategies in the absence of a therapist. Volunteer young people from the target age groups were included in the development process, providing feedback on the look, feel, and functionality of the key added components (eg, relaxation room and exposure hierarchy tool) through two iterations. In addition, an expert advisory panel comprising the research team, two expert advisors, stakeholder representatives, and two youth advisors provided feedback throughout the development and delivery process.

In addition to the removal of any therapist contact, the following changes were made to the existing intervention. First, we created a new infrastructure to surround the existing program and provide clearer navigation to resources for users. This included additional home pages, new graphics, demonstration videos, frequently asked questions, and dedicated sections for resources and key program components (eg, relaxation room and exposure section). Second, we integrated relaxation activities into the program via the dedicated relaxation room, where users could live stream relaxation or download relaxation recordings (or transcripts) to add to their music libraries. Third, the exposure hierarchy (BRAVE ladder) was integrated into the new home page infrastructure in a dedicated section and additional tools were created to ensure that young participants were able to build their exposure hierarchy effectively without a therapist. This included easy creation of steps and rewards, the ability to move steps around, and the capacity to check-off steps when complete. Fourth, in addition to the already included automatic email reminders and session completion messages, the self-help program also included automatic alert messages that were sent to the users if they reported anxiety scores in the clinical range. Finally, we integrated a self-registration system into the program that required young people to register for the program and provide a contact email address and name or pseudonym. Registration also included the provision of consent and an explanation of the monitoring and reporting features included in the program. All users were able to first trial the program in a 20-minute guest access before registering. Young people were able to register for and complete the program on their own, via a personal computer or mobile device. Program sessions were conducted in a prescribed sequence, although participants could progress at their own pace. Furthermore, automatic email reminders were sent if participants had not completed the next session within 1 week.

Measures

In this study, we aimed to disseminate the intervention as widely as possible on a national scale; thus, we implemented assessment procedures conducive to this goal. The advisory committee identified administering an exhaustive assessment battery as a potential deterrent for young people registering and completing the program in this self-directed manner. Thus, to minimize participant burden and the potential barriers to participation, we administered a limited assessment battery, including one brief measure of anxiety, along with basic demographic characteristics and a brief satisfaction scale.

Figure 1. The progression of participants through the program. CAS-8: Children's Anxiety Scale, 8-item.



Demographics

Demographic data (eg, age, gender, and postcode) were collected when participants created their account to access the program. Postcodes were categorized according to the Australian Standard Geographic Classification system [17] and coded into Major City, Inner Regional, Outer Regional, Remote, and Very Remote. For the purpose of categorical analysis, these categories were coded into Major Cities and Outside Major Cities.

Anxiety

We measured anxiety symptom severity using CAS-8 [18], an 8-item scale adapted from the Spence Children's Anxiety Scale [19], assessing child anxiety symptoms on a 4-point scale (0=*Never*, 3=*Always*). The CAS-8 has demonstrated good reliability and provides population-level, gender-standardized norms for comparison [18]. Scores of ≥84th percentile (ie, above a *T*-score of 60: CAS-8 score ≥10 for males and ≥12 for females) are considered indicative of elevated anxiety, while scores of ≥94th percentile (ie, above a *T*-score of 65: CAS-8 score ≥13 for males and ≥16 for females) are considered representative of clinical levels of anxiety. In this study, participants completed the CAS-8 prior to beginning the program (baseline) and at the beginning of Sessions 4, 7, and 10 (ie, after completion of Sessions 3, 6, and 9). In addition, the CAS-8 was integrated into the program such that it was conducted before the participant could progress with the session. The internal consistency of the CAS-8 for data collected throughout this program was 0.85.

Adherence and Satisfaction

Session completion (adherence) was operationalized as the number of program sessions completed and was automatically recorded by the program. We measured both satisfaction and acceptability via a 5-item scale based on a satisfaction questionnaire administered in previous trials of the BRAVE Program [10,11]. Satisfaction data were measured at the same time points as the CAS-8 and were examined based on the responses to the final (latest) satisfaction assessment completed by each participant. In addition, participants were required to respond to items assessing whether they would tell a friend about the program (Item 1), how helpful the program was (Item 2), how happy they were with the program (Item 3), how much the program helped to reduce their anxiety (Item 4), and overall judgment of the program (Item 5). Responses to the 5 items were provided on a 5-point Likert scale, with responses for item 1 scored as 1=*Definitely Not*, 3=*Maybe*, and 5=*Definitely Yes*; responses for items 2, 3, and 4 scored as 1=*Not at all*, 3=*Quite a bit*, and 5=*Very Much*; and responses for item 5 scored as 1=*Very Bad*, 3=*Okay*, and 5=*Very Good*. We calculated the mean item and mean total satisfaction scores. An additional final item was included as a free-text, qualitative item asking the participant to comment on anything else about the program. A sample of responses to this question is provided in the Results.

Safety Alerts

Given the self-help nature of this intervention, the program was designed to incorporate checks on anxiety, alerts regarding participants who were experiencing high levels of anxiety, and the provision of appropriate referral information. In this study, any child or adolescent scoring in the clinical ranges of anxiety at any assessment point was sent an automatic message that alerted the person to his or her high score and encouraged him or her to seek further support from additional sources, including family, friends, and professional services. The message also included contact details for crisis lines and services. We calculated the proportion of children and adolescents receiving email alerts at the four different assessment points.

Statistical Analysis

We used IBM Statistics 24 and MPlus 8 for statistical analyses. Descriptive data for satisfaction item and total means were evaluated and presented for the total sample as well as child and adolescent subsamples. Examples of feedback are provided as well. Furthermore, descriptive data for program adherence (number of sessions completed) were evaluated and reported for the total sample as well as according to child and adolescent samples. In terms of safety alerts, the proportion of participants receiving email alerts at each time point was calculated. Baseline differences in anxiety severity, age, gender, and geographic location between the participants who completed less than three sessions and those who completed three or more sessions were examined using *t* tests and chi-square tests. Furthermore, the relationship between program compliance and CAS-8 scores at the final program attendance was expressed as a Pearson product-moment correlation coefficient (*r*). Change in anxiety was analyzed in three ways following recommendations for determining therapeutic changes in child and adolescent populations [20]. Given the lack of a control comparison condition, providing multiple inferences of the data through different means allowed checks for common patterns in results and increased confidence in the results observed.

First, we examined the mean change in raw anxiety scores across the program. Following the procedure of Rickwood et al [21] in their implementation evaluation of the headspace service, analyses were first conducted with treatment "completers," ie, those who provided data at the relevant time points. Therefore, to determine changes in anxiety scores from baseline to Session 4, baseline to Session 7, and baseline to Session 10, separate repeated-measures analyses of variance (ANOVA) were conducted with individuals who had completed the sessions and the assessment measures up to that point. Given that users could not be followed up after they left the program, the last data point provided within the program represents a participant's final data point. Analyses were also conducted based on the participants' baseline and final assessment points (last assessment completed) to provide an overview of outcomes from the open-access program irrespective of the amount of treatment completed. For the majority of the sample, the final score was representative of approximately three sessions completed. Furthermore, a post-hoc power analysis revealed that the "completer" sample size for children (n=532) and adolescent (n=563) groups provided the power of 1.00 in detecting the within-subjects effect over four time points.

Given the large amount of missing data from participants who failed to provide data at all four assessment points, we also analyzed the full sample through latent growth curve modeling (LGCM) to confirm the findings of the completer analysis. LGCM is accepted as a suitable framework for use in the evaluation of efficacy in psychological interventions [22] and allows determination of whether the temporal trajectory from the baseline to Session 10 is significant, using data from all participants rather than just the treatment completers [22]. Time was specified as linear over the active treatment phase (baseline, Session 4, Session 7, and Session 10) in the growth models, and residual variances were held equal across time. Furthermore, models were fitted with a full-information maximum likelihood estimator using the Mplus 8 program [23]. We examined anxiety symptom trajectories separately for child and adolescent samples, in line with the repeated-measures ANOVA. For both the ANOVA and growth curve models, results were converted to standardized effect sizes (Cohen *d*).

The second method for analyzing changes in anxiety included utilization of the Reliable Change Index (RCI) [24]. RCI is a psychometric criterion that evaluates whether an individual participant changes sufficiently on a target measure (eg, CAS-8) over time (eg, from baseline to Session 4) and whether this change can be considered statistically significantly greater than the difference that might have been expected due to measurement error or unreliability [24]. In addition, RCI assesses whether the difference between two scores is more than a set level, determined by the product of the instrument's SD and reliability [24]. Changes in scores can subsequently be categorized as "reliable improvement," "no improvement," or "reliable deterioration." Given the lack of a control comparison condition in this study, RCI affords an opportunity to provide more rigorous analysis of the data and to illustrate statistically

reliable change at the individual level for those participants who provided data at more than one time point. In this study, we calculated reliable change scores for each participant using the CAS-8 and presented the proportion of participants demonstrating reliable improvement, no improvement, or deterioration for each of the time contrasts (baseline to Session 4, baseline to Session 7, and baseline to Session 10) outlined above. Furthermore, RCI was estimated to be equivalent to a 4-point change for males and a 5-point change for females using reliability coefficients and gender-standardized norms for the CAS-8 based on Australian school-aged youth [18].

The third method of assessing change in anxiety was through examination of the proportion of individual youth cases crossing the clinical threshold [20] into recovery at a group level. The clinical threshold was deemed as a CAS-8 score ≥94th percentile (ie, *T*-score of ≥65) based on a large-scale community sample [18]. Of all participants, 51.32% (2271/4425; 632/1473 [42.91%] children and 1639/2952 [55.52%] adolescents) were categorized as being above the clinical threshold on the CAS-8 at enrollment. The remaining 48.68% (2154/4425) of the sample met criteria for "elevated" but not "clinical" levels of anxiety. We, therefore, also examined the proportion of youth crossing from the "elevated" threshold (84th percentile, *T*=60) to the "nonelevated" range. All analyses are presented separately for child and adolescent program users.

Results

Program Satisfaction

Using the last provided satisfaction scores for each participant, the mean total satisfaction rating was 17.72 (SD 5.16) out of a maximum 25. The mean satisfaction ratings for the individual items are provided in Figure 2 for all participants and across children and adolescents.

In terms of the open feedback item, Figure 3 presents a list of comments provided by a snapshot of participants.

Program Adherence

As indicated in Figure 1, only 24.75% (1095/4425) of the participants who demonstrated elevated anxiety at registration provided data at the second assessment point at Session 4 (and, thus, had completed at least three sessions). Participants who completed three or more sessions were younger (mean 11.90 years, SD 2.88) than those completing less than three sessions (mean 13.40 years, SD 2.89; $F_{1,4423}$=252.68; *P*<.001). Furthermore, the former showed lower baseline anxiety severity (mean 14.86, SD 3.26), than the latter (mean 15.42, SD 3.40; $F_{1,4423}$=26.02; *P*<.001). In addition, the former were more likely to be females (827/1341, 61.67%) than males (495/1341, 36.91%; $\chi^2_{1,4344}$=22.4; *P*<.001) and other gender (19/1341, 1.41%) and were more likely to reside in major cities (741/1282, 57.80%) than in nonmetropolitan areas (541/1282, 42.20%; $\chi^2_{1,4184}$=6.8; *P*=.009).

Of note, a large proportion of registered participants (958/4425, 21.65%) did not go on to complete the first session. A further 48.05% (2126/4425) of registered participants completed only one or two sessions of the program. However, 30.31% (1341/4425) participants went on to complete three or more sessions, with 1095 of these providing assessment data for at least two time points. The average number of sessions completed for all registered participants (including those who did not start the program) was 2.21 (SD 2.44). Figure 4 provides a visual summary of how many sessions were completed by all participants who had registered for the program, including a breakdown of sessions completed by children and adolescents.

Figure 2. The mean satisfaction ratings for individual satisfaction items.

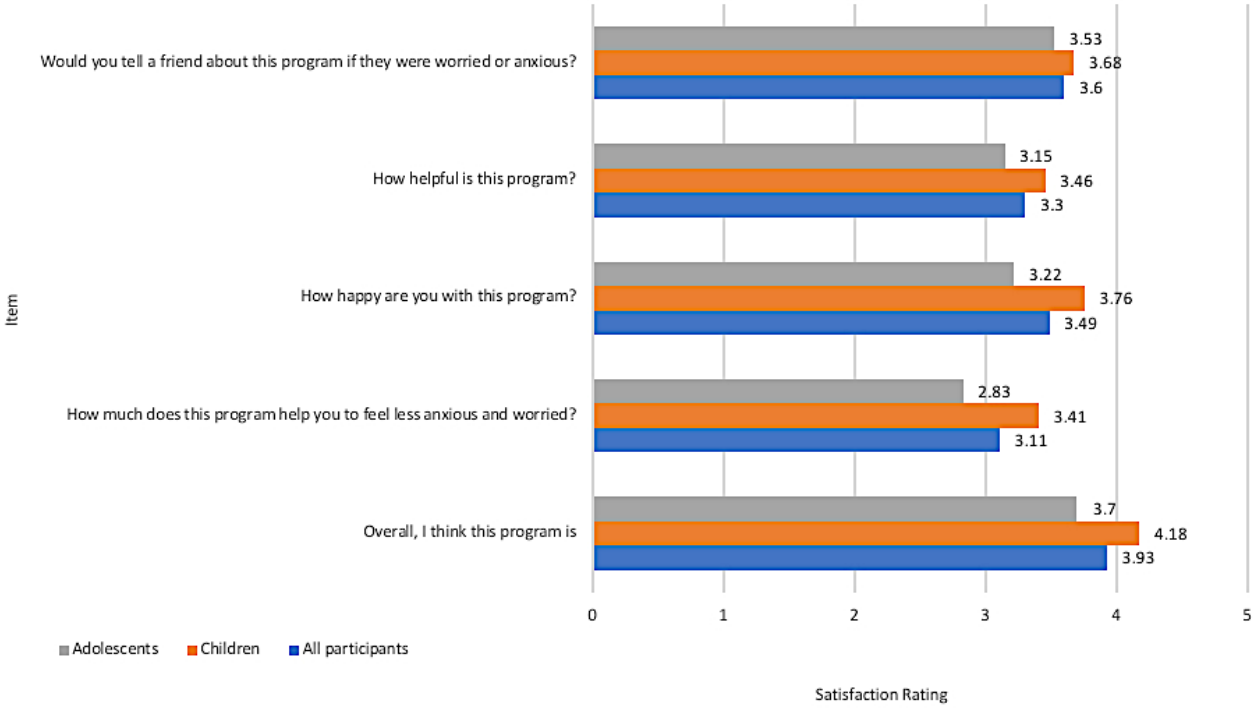


Figure 3. Feedback comments from participants.

Feedback from participants	
"I love how the program checks your understanding of concepts by giving you fun quizzes and extra challenges. I appreciate how I receive emails to remind me to practice the skills I have learnt."	"When I first started I didn't see the point and I didn't get why they wanted me to do it but now I understand and get why it helps a lot and I'm grateful that anyone even cares..."
"It isn't boring and it's quite fun but learning fun"	"I know how to calm myself down now"
"I like the way they teach us step by step how to do muscle relation"	"It needs to be personalized"
"This program is awesome and is helping me in class and in trampolining"	"it's good for people who never tell anybody things"
"It's a fun and helpful program"	"Too much writing, not enough pictures"
"I always tell my mum or dad my problems, but writing them down has helped"	"Sometimes it feels like school work with the regular updates to continue through with the program"
"So far this program is really good, I'm personally gaining a lot from it. However i feel like some components are a little childish, but thats okay."	"The program may work for some and not for others, I have found it helping me a little bit but then again I am only up to part 4. So I hope it will help me feel less anxious around people at school!"
"Is there an app I can use instead of the internet?"	"When I go on this program I get braver and braver every time"
"That it has given me the confidence I need to do things I find hard"	"Lessons are too long"

For those participants who provided at least two data points (and, therefore, had completed at least three sessions), the average number of sessions completed was 5.69 (SD 2.35) out of 10. On average, child users completed 5.77 sessions (SD 2.45), while adolescent users completed 5.62 sessions (SD 2.25). A significant negative correlation was evident between the number of sessions completed and final CAS-8 scores (r=−.17; P=.003), such that higher session completion correlated with the lower final anxiety severity.

Mean Changes in Anxiety

Changes in the CAS-8 score over time are provided for the child program participants in Figure 5 and adolescent program participants in Figure 6 (completer sample). These graphs plot the mean CAS-8 scores at the four time points for which anxiety was assessed and relate to those individuals who were retained in the program up to and including the session in which the assessment was completed. Thus, the number of participants present at each assessment point decreased as the number of completed sessions increased, as described above.

Figure 4. A visual summary of how many of the registered participants (N=4425) completed how many sessions of BRAVE Self-Help during the 20-week period (including participants who only provided one assessment point).

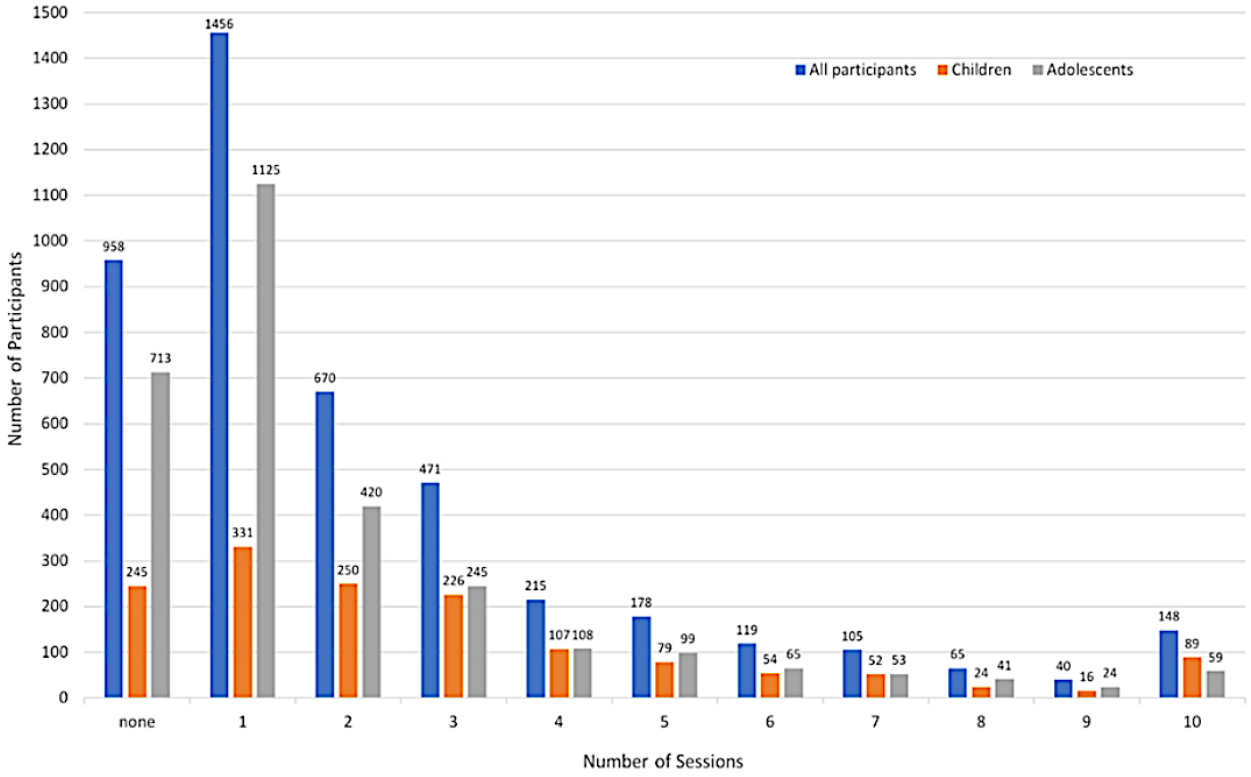


Figure 5. Changes in the mean anxiety scores according to the number of sessions completed for child program users. CAS-8: Children's Anxiety Scale, 8-item.

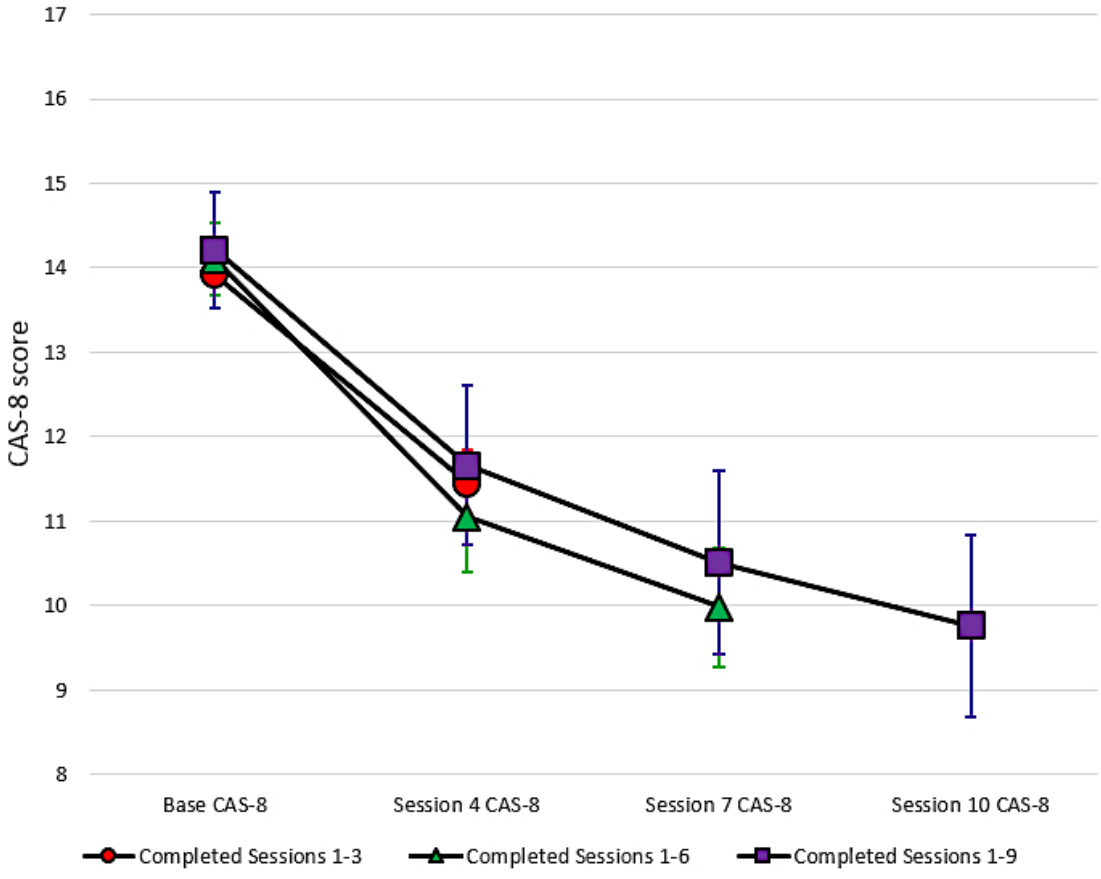
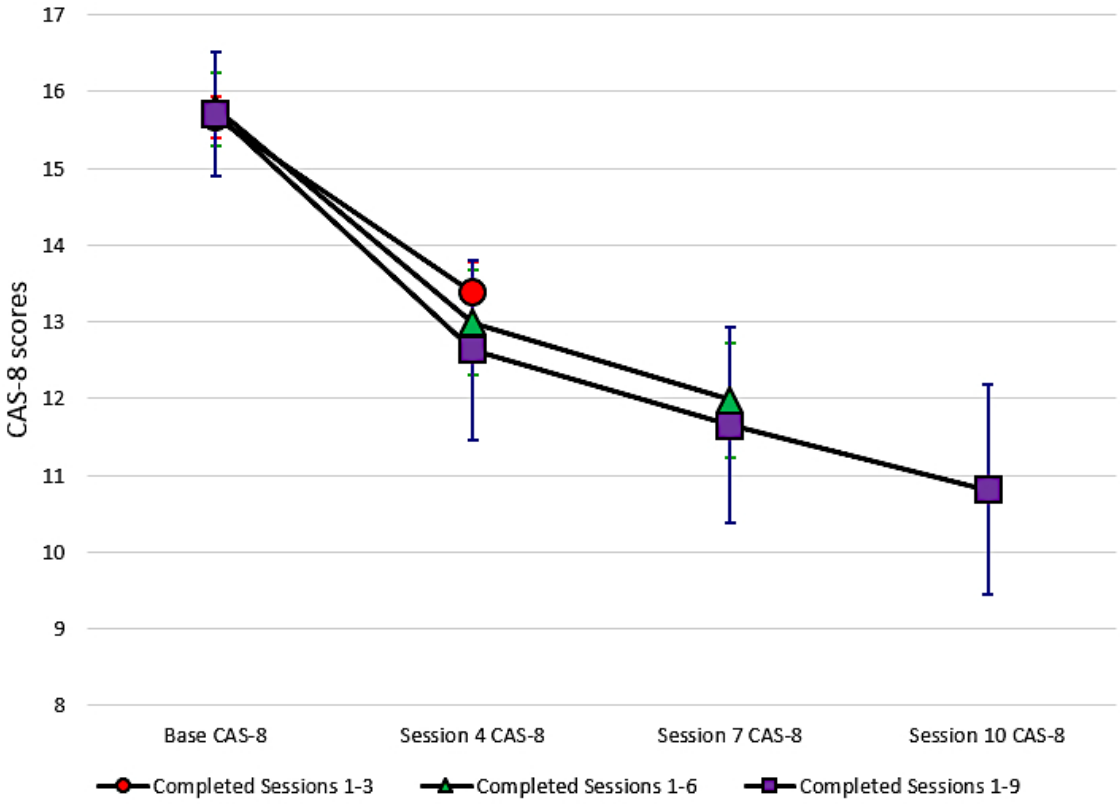


Figure 6. Changes in the mean anxiety scores according to the number of sessions completed for adolescent program users. CAS-8: Children's Anxiety Scale, 8-item.



The mean CAS-8 scores over time and the results of ANOVAs and effect sizes are presented in Multimedia Appendix 1. We observed significant reductions in anxiety from the baseline to Session 4, baseline to Session 7, and baseline to Session 10 for both children and adolescents. For users who completed six or more sessions, we noted an average 4-point improvement in CAS-8 scores (Cohen *d*=0.88, children; Cohen *d*=0.81, adolescents), indicating a moderate to large effect size. The mean change was slightly less from the baseline to Session 4, although still significant, and was greater for those who completed the program up to the final Session 10 (see Multimedia Appendix 1). Figure 7 shows the change in CAS-8 scores for all participants from the baseline to their final data point before ending their engagement with the program. Results show that irrespective of the number of sessions completed, young people showed a statistically significant average decrease of around 3 points on the CAS-8 for both the child (Cohen *d*=0.66) and adolescent (Cohen *d*=0.65) programs.

The results for LGCMs confirm the findings of the completer sample and demonstrate that anxiety decreased significantly over time. Estimated means, SDs, and effect sizes for the CAS-8 in the child and adolescent samples are provided in Table 1. Linear growth curve models were estimated to evaluate the effect of time from the baseline to Session 10 on this anxiety outcome measure. For children, a statistically significant decrease was noted in anxiety from the baseline to Session 10 (*B*=−1.95; beta=−1.13; standard error [SE]=0.13; *P*<.001). In the adolescent sample, likewise, a statistically significant decrease was noted in anxiety from the baseline to Session 10 (*B*=−1.89; beta=−1.00; SE=0.13; *P*<0.001).

Reliable Change

The proportion of child and adolescent users demonstrating statistically reliable (gender-adjusted) improvement, deterioration, or no change according to sessions completed is presented in Multimedia Appendix 2. The percentage of youth showing statistically reliable improvement increased as the number of sessions completed increased, with no significant differences between the proportions of children and adolescents showing change. Across the entire sample, for those who completed nine sessions (assessment completed at the beginning of Session 10), 54.6% (89/163) demonstrated statistically reliable improvement on the CAS-8, 40.5% (66/163) showed no statistically reliable change, and 4.9% (8/163) showed deterioration. For the entire sample, irrespective of the number of sessions completed, by their final recorded assessment, 35.62% (390/1095) showed reliable improvement in anxiety, 59.91% (656/1095) showed no statistically reliable change, and only 4.47% (49/1095) showed deterioration. Importantly, for child participants, only a very small proportion showed deterioration. Specifically, 4.32% (23/532), 3.63% (7/193), and 7.70% (7/91) of children showed deterioration after three, six, and nine sessions, respectively. This figure was even lower for adolescents, with only 3.20% (18/563), 4.88% (10/205), and 1.39% (1/72) showing deterioration after three, six, and nine sessions, respectively.

Figure 7. Changes in the mean anxiety scores from a user's baseline to final CAS-8 score. CAS-8: Children's Anxiety Scale, 8-item.

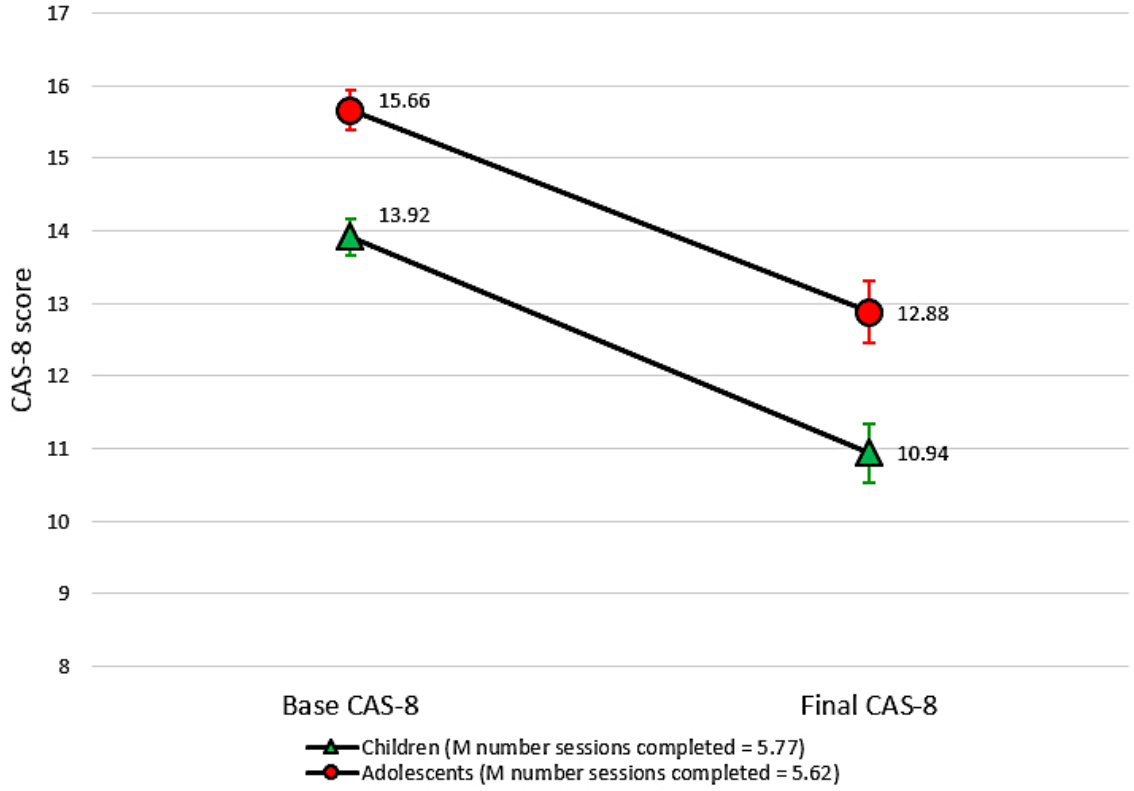


Table 1. Estimated Children Anxiety Scale, 8-item, means, SDs, and effect sizes from growth curve analyses.

Group	Baseline, mean (SD)	Session 4		Session 7		Session 10	
		Mean (SD)	Cohen *d*[a]	Mean (SD)	Cohen *d*[b]	Mean (SD)	Cohen *d*[c]
Children	14.09 (3.06)	11.55 (4.59)	0.60	10.24 (4.96)	0.82	9.57 (5.17)	0.86
Adolescents	15.83 (3.36)	13.51 (4.70)	0.60	12.42 (5.47)	0.74	11.53(5.77)	0.89

[a]Effect size from the baseline to Session 4.

[b]Effect size from the baseline to Session 7.

[c]Effect size from the baseline to Session 10.

Proportion Crossing the Elevated and Clinical Thresholds

The proportion of child and adolescent users crossing the "elevated" and "clinical" thresholds at the different time points (and according to sessions completed) is provided in Multimedia Appendix 3. The change from elevated or clinical anxiety status into nonelevated anxiety status is an indicator of recovery. Similarly, the change from being above the clinical threshold to elevated (but not clinical) is an indicator of response (clinically meaningful improvement). The proportion of youth demonstrating recovery or response on these indicators increased as the number of sessions completed increased. For those who completed six sessions (Session 7 assessment), 53.0% (211/398) crossed from the "elevated" to "nonelevated" range (ie, demonstrated recovery); this increased slightly to 57.7% (94/163) for those who completed nine sessions. Across all users, 53.88% (590/1095) showed a reduction from the "elevated" anxiety range into the "nonelevated" range from the baseline to their final CAS-8 score before ending their engagement with the program, although a significantly greater proportion of children (325/532, 61.1%) than of adolescents (265/563, 47.1%) showed this change ($\chi^2_{1,1095}$=21.64; *P*<.001).

For those who initially demonstrated "clinical" levels of anxiety, 59.1% (309/523) demonstrated an improvement into the "nonclinical" range by their final recorded score. We observed no differences in the proportion of children (134/217, 61.8%) and adolescents (175/306, 57.2%) achieving this level of improvement (χ^2=1.09; *P*=.30). As is evident in Multimedia Appendix 3, of those who were initially in the "clinical range" of anxiety and who completed at least nine sessions of the program, 52 of 74 youth (26/37 children; 26/37 adolescents) no longer experienced clinical levels of anxiety. As demonstrated in Multimedia Appendix 4, of those participants demonstrating clinical anxiety at the baseline, 34 of 74 (17/37 children; 17/37 adolescents) reduced to the normal range of anxiety (recovered) and 18 of 74 (9/37 children; 9/37 adolescents) reduced to the elevated range (responded).

Safety Alerts

The proportion of young people receiving email alerts for the presence of clinical-level anxiety was higher among adolescents than among children across all four time points and showed reduction over time for both groups. At the baseline (registration), 59.96% (1770/2952) of adolescents and 38.15% (562/1473) of children received email alerts, with this difference being statistically significant ($\chi^2_{1,4425}$=187.5; P<.001). For adolescents, this number dropped to 39.8% (224/563) for those completing three sessions, 31.2% (64/205) for six sessions, and 29% (21/72) for nine sessions. For children, only 21.1% (112/532) received email alerts after completing three sessions, 14.5% (28/193) after six sessions, and 16% (15/91) after nine sessions.

Discussion

Study Objectives

This study reports on the feasibility and acceptability of a free, self-help iCBT intervention, offered nationally in Australia, and highlights the potential benefits, yet significant challenges evident, with this type of service delivery. The BRAVE Self-Help initiative was designed to provide an evidence-based intervention to anxious Australian children and adolescents while minimizing user burden and barriers to receipt of treatment (eg, cost, accessibility, stigma, and privacy). This study examined the feasibility and acceptability of this approach by evaluating satisfaction with, and adherence to, the program as well as changes in anxiety symptoms. Although there was no control comparison condition or trial methodology, the methodological approach to analysis was comprehensive, with data being collected on multiple occasions and the effects assessed using multiple methods of evaluating changes in anxiety. Furthermore, both satisfaction and adherence data provided further information about the acceptability of the open-access intervention.

Intervention Acceptability

The BRAVE Self-Help program is extremely comprehensive as all specialist techniques are incorporated into the content and prominent examples and opportunities for skill rehearsal are integrated within and between sessions. In addition, the content is interactive, engaging, and age appropriate, and our previous research has demonstrated high consumer acceptability and satisfaction [10,11]. Thus, it was not surprising to find moderate to high satisfaction ratings reported by children and adolescents participating in the BRAVE Self-Help program. Slightly lower satisfaction ratings were reported by adolescents for all items, although the ratings remained moderate overall. As a group, participants were happy with the program, rated the overall program highly, and would refer the program to a friend if he or she experienced anxiety. Interestingly, satisfaction reported by users of this self-help version of the program was highly similar and in some cases, higher than that reported by youth participating in the therapist-assisted version of the program in our previous randomized controlled trials (RCTs; child program, mean satisfaction rating 3.6/5) [10] (adolescent program, mean satisfaction rating 3.53/5) [11]. Thus, satisfaction was not diminished when delivered as a self-help, open-access program with widespread dissemination.

Qualitative feedback indicated that many users were able to obtain benefits from the program, and potential improvements were noted by others. In particular, the intervention might benefit from accompanying app-based features, either for the entire program or for specific intervention components such as exposure. Furthermore, when delivered in this self-directed format, sessions might need further refinement to decrease the length, increase the use of videos and acceptable graphics, and reduce monitoring and reminder systems (unless requested, or demonstrated as being important in stimulating engagement). Interestingly, the need for sessions to be personalized was raised, despite the intervention being run as a self-help program without any professional contact. Thus, there are opportunities to implement innovative technology-based methods for achieving treatment personalization (eg, utilizing algorithms to present personalized treatment content or messages based on previous session responses) in such open-access, self-help online interventions.

Despite positive feedback and satisfaction with the program, there was a noticeable variation in the degree of program adherence across users. The fact that around 21% (958/4425) of participants did not go on to complete any of the sessions and a further 48% (2126/4425) completed only two sessions or less indicates that the program was either not acceptable or potentially not useful to a substantial proportion of people. These rates of session progression are not dissimilar to those found in other large-scale, open-access eHealth interventions such as MoodGym (in a sample of 82,159 participants, 63% completed no modules, 27% completed only 1 module, and 10% completed 2 or more modules out of 5 modules) [25] or even in face-to-face clinic service delivery contexts such as headspace (mean sessions attended 4.1; 49% of patients completing two or less sessions) [21]. Furthermore, in the only existing implementation trial for childhood anxiety, Jolstedt et al [15] reported an average adherence of 6 of 12 modules, with only half of the sample reaching module 4, despite the presence of therapist assistance and parent involvement.

In all likelihood, these results may suggest that services that are widely or publicly available attract users from varying contexts and backgrounds, who will subsequently engage very differently with the programs than those in strictly controlled RCTs. With respect to the BRAVE Self-Help initiative, given the significant change observed in users completing only three sessions, it is possible that several participants engaged in the program until they obtained the benefit they needed, which might have occurred early on in the program. For others, such as those who did not commence the program at all, it is possible that their expectations were not aligned with what the program had to offer and that this only became apparent after registering. For others still, it may be possible that a self-directed program with little guidance was simply not enough to sustain engagement. In this study, we were unable to determine the reasons for nonadherence, and this will be important for future studies. Nonetheless, while the program was acceptable, the low rates of adherence suggest that such approaches will not be sufficient for or acceptable to all.

Changes in Anxiety

The results of this study demonstrate that self-reported anxiety decreased significantly over time, with effects being greater as the number of sessions completed by youth increased. These findings were confirmed through the completer analyses (N=1095) and growth curve modeling, which utilized all eligible participants registering for the program (N=4425). Improvements were evident for both children and adolescents, although adolescents showed slightly less improvement on some outcome indicators. Based on the assessment point at the beginning of Session 4, results showed that around 43% of children and around one-third of adolescents had recovered (no longer experienced elevated anxiety). For those with clinical levels of anxiety, over half (57%) of children and just under half (45%) of adolescents were no longer in the clinical range after completing three sessions (ie, showed response). Importantly, even greater reductions were evident for child and adolescent users who completed more sessions, such that of those who completed nine sessions, over two-thirds (70%) were no longer in the clinical range (ie, showed response) and almost half (47%) no longer demonstrated elevated anxiety (ie, showed recovery). Furthermore, of all participants who completed nine sessions, around half (54%) achieved statistically reliable change as indicated by RCI. Thus, we observed substantial reductions in anxiety across multiple measures among users of a freely available, evidence-based online intervention.

Although not directly comparable, as diagnostic status was not determined by clinical interview in this study, the results can be compared with those of the previous RCTs of the therapist-assisted BRAVE-ONLINE program. The demonstrated effect sizes for reductions in anxiety from the baseline to Session 9 in this study (Cohen *d*=0.83-1.01) are similar to the effect sizes observed in previous RCTs (Cohen *d*=0.91-1.23 for the full Spence Children's Anxiety Scale) [10,11]. Furthermore, the proportion of young people with clinical levels of anxiety at the baseline who no longer reported anxiety (47%) after participating in BRAVE Self-Help is similar to the proportion of youth no longer meeting diagnostic criteria after 12 weeks in the above-mentioned RCTs (between 30% and 37%). However, it is important to note that different (briefer) outcome measures were utilized in this study compared with the comprehensive diagnostic assessments obtained in RCTs. Thus the results of this study are not directly comparable and should be interpreted with caution.

The effect sizes for BRAVE Self-Help presented in this study are somewhat smaller than those of a pilot implementation study (N=20) conducted very recently by Jolstedt et al [15]. In their study, the authors reported an effect size of Cohen *d*=1.22 on the full Spence Children's Anxiety Scale from pre- to posttreatment, although it should be noted that this intervention was (1) delivered with substantial therapist support (regular asynchronous contact through messages, comments on activities, and phone calls), (2) required both parents and children to complete sessions before subsequent sessions were unlocked, and (3) encouraged families to log in at least twice a week to respond to therapists [15]. Thus, BRAVE Self-Help is a lower-intensity intervention than the one described by Jolstedt et al, yet with only somewhat lesser effects.

Implications

It is often argued that low-intensity interventions such as iCBT programs are only effective with professional support. However, the results of this study demonstrate that self-help online interventions may be effective for many young people if they complete at least three sessions of the program. Indeed, the improvements demonstrated by participants in this open-access, self-help program are significant and reveal that a meaningful change is feasible without therapist support. The finding that a substantial proportion of young people can achieve clinically meaningful improvements through such an intervention in a self-help format has significant implications for models of service delivery. In addition, substantial therapist time and cost savings may be afforded by such self-help interventions, contributing to an increase in the overall efficiency of youth mental health services. Furthermore, providing evidence-based services via online self-help may reach more young people at a far lower cost than face-to-face service delivery models. It would seem that if young people who would benefit from self-help programs are accurately identified, more costly resources (eg, face-to-face therapist sessions) could be reserved for those young people who need them the most.

Another implication of the findings relates to the treatment dose or magnitude of change in early sessions. Specifically, it is worth noting that the highest magnitude of change was evident following the completion of the first six sessions. Thus, the results of this study demonstrate that even a small dose of self-help treatment may be effective for some young people and would perhaps bring about a change equivalent to that from an extended program. These findings are also consistent with those of Chu et al [26] who demonstrated a nonlinear symptom trajectory for youth engaging in face-to-face CBT; in fact, they reported that participants tended to show a rapid response over the first six sessions, with changes tending to taper off thereafter. Therefore, if used as the first step in a stepped-care service provision model, users may potentially require only six sessions or fewer. This makes sense given that the first six sessions of the program contain the specialist CBT skills and skill rehearsal, with Sessions 7-10 targeting practice with exposure tasks and relapse prevention. Therefore, not only do the results of this study support the benefits of online self-help interventions but also assist in identifying the potential trajectory of a symptom change or the "ideal dose" of such low-intensity programs.

Limitations and Future Research

This open dissemination study has some design limitations. There was no control group against which to determine whether changes simply reflect spontaneous recovery, regression to the mean, or nonspecific intervention effects. In addition, there was no initial interview to confirm the clinical diagnosis, instead a single informant and a single measure was relied upon. Although this measure is unable to provide a comprehensive overview of anxious symptomatology type and intensity, it does allow for examination of the magnitude of anxiety symptom change, which was the objective of this study. Furthermore, it was necessary to minimize participant burden in an open trial of this nature where participation was completely voluntary and self-sought.

We also note that some analyses in this study were limited to those individuals who completed at least two assessment points and, thus, had completed at least three program sessions. While this setting provides a fair evaluation of the outcome for those who completed at least a proportion of the program, it does not consider those who initially enrolled and decided not to continue at all with the program. Perhaps, such young people were not ready to participate in active treatment or they might require alternative treatment modalities. In fact, in this study, those who completed less than three sessions were more likely to reside in non-metropolitan areas, were older and had a higher anxiety severity. It is possible that low-intensity iCBT programs may be more accessible or preferred by youth who live in major cities. Also, younger children may be more likely than adolescents to complete iCBT programs with parental assistance, which may bring about higher adherence to the program. Furthermore, the finding that youth with lower levels of anxiety adhered more to the program is consistent with the objectives of low-intensity iCBT interventions and supports the notion that iCBT for youth could be useful as a first step in intervention.

In terms of sample representativeness, all participants in this study demonstrated elevated anxiety > 84th percentile on the CAS-8. Thus, all participants were experiencing anxiety at a level that was interfering with their lives; however, only half of the participants demonstrated "clinical" levels of anxiety (>94th percentile). Thus, the sample was somewhat less severe than those examined in previous RCTs of iCBT for child anxiety [10,11]. Participants in this study included both children and adolescents, with similar mean ages to other trials on child anxiety. However, there was a slightly higher proportion of females (61%) in this study than in previous trials that had a more even gender distribution [10,11]. Finally, although the BRAVE Program does offer parent modules, in this open-access, anonymous delivery model, we were unable to link parent and child accounts and, thus, could not determine the impact of parental involvement.

Despite its limitations, the findings of this study are encouraging and certainly justify further research in an RCT or similar design to confirm the effectiveness of this type of intervention. Historically, there has been a lack of formalized approaches to the dissemination of iCBT interventions [14], yet there are some recent design recommendations that may be appropriate. To address the challenges of conducting controlled trials in real-world implementation settings, it may be necessary to conduct a stepped-wedge cluster design, which is increasingly recommended for evaluation of service delivery interventions [27]. Such research should also consider the inclusion of clinician-, parent-, or teacher-informant reports to provide a more comprehensive diagnostic information. Furthermore, future research should incorporate more frequent assessments of anxiety to enable the examination of response trajectories according to baseline individual and clinical factors. Such analyses will assist in defining the ideal treatment dose as well as predictors of nonresponse and have the potential to provide evidence that could inform the design and implementation of "stepped-care" approaches. Finally, there is a pressing need for health economic evaluations to determine the relative cost and health benefits of online, self-help approaches compared with therapist-supported, online programs and face-to-face services.

Conclusion

This online, self-help program for anxiety has the capacity to reach greater numbers of young people compared with programs that require therapist contact. The results of this open trial demonstrate moderate to high program acceptability when delivered in this way and show its potential feasibility in bringing about clinically and statistically meaningful reductions in anxiety for children and adolescents. Greater reductions were evident for those who completed more sessions, although significant improvements were most evident in the first six sessions. Overall, self-help iCBT is a potentially feasible and acceptable approach for delivering evidence-based interventions through a public health delivery model.

Acknowledgments

The authors wish to acknowledge the funders of this project, *beyondblue*; the many participants who participated in this research; the research assistants working on the project; and the members of the advisory panel (Professors David Kavanagh, Jane Burns, Britt Klein, *beyondblue*, and *blueVoices* youth representatives) who provided guidance on the overall implementation of the project.

Conflicts of Interest

SM, SHS, and CLD acknowledge that although intellectual property for BRAVE-ONLINE is owned by UniQuest/the University of Queensland, they may potentially benefit from royalties related to the program.

Multimedia Appendix 1

A summary of analysis of variance (ANOVA) and effect sizes.

[PDF File (Adobe PDF File), 28KB - jmir_v20i7e234_app1.pdf]

Multimedia Appendix 2

A reliable change over time.

[PDF File (Adobe PDF File), 16KB - jmir_v20i7e234_app2.pdf]

Multimedia Appendix 3

The proportion of participants crossing the "elevated" and "clinical" thresholds.

[PDF File (Adobe PDF File), 49KB - jmir_v20i7e234_app3.pdf]

Multimedia Appendix 4

The proportion of clinical participants moving from "clinical" to "elevated" and "normal" anxiety levels.

[PDF File (Adobe PDF File), 48KB - jmir_v20i7e234_app4.pdf]

References

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
2. Lawrence D, Johnson S, Hafekost J, Boterhoven DHK, Sawyer M, Ainley J. Report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing. Department of Health, Canberra. 2015. The Mental Health of Children and Adolescents URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/9DA8CA21306FE6EDCA257E2700016945/$File/child2.pdf [accessed 2018-06-14] [WebCite Cache ID 70A25qGo1]
3. Warwick H, Reardon T, Cooper P, Murayama K, Reynolds S, Wilson C, et al. Complete recovery from anxiety disorders following Cognitive Behavior Therapy in children and adolescents: A meta-analysis. Clin Psychol Rev 2017 Mar;52:77-91. [doi: 10.1016/j.cpr.2016.12.002] [Medline: 28040627]
4. Beesdo K, Knappe S, Pine DS. Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V. Psychiatr Clin North Am 2009 Sep;32(3):483-524 [FREE Full text] [doi: 10.1016/j.psc.2009.06.002] [Medline: 19716988]
5. James AC, James G, Cowdrey FA, Soler A, Choke A. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev 2013;6:CD004690. [doi: 10.1002/14651858.CD004690.pub3] [Medline: 23733328]
6. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
7. Christensen H, Reynolds J, Griffiths KM. The use of e-health applications for anxiety and depression in young people: challenges and solutions. Early Interv Psychiatry 2011 Feb;5 Suppl 1:58-62. [doi: 10.1111/j.1751-7893.2010.00242.x] [Medline: 21208393]
8. Australian Government. Australian Government Response to Contributing Lives, Thriving Communities - Review of Mental Health Programmes and Services. Canberra: Department of Health; 2015. URL: http://www.health.gov.au/internet/main/publishing.nsf/content/0DBEF2D78F7CB9E7CA257F07001ACC6D/$File/response.pdf [accessed 2017-10-10] [WebCite Cache ID 6u7Vd9Nf2]
9. Donovan C, March S. Computer-based treatment programs for youth anxiety: A systematic review. Psychopathol Rev 2014 Jul 28;1(1):130-156. [doi: 10.5127/pr.033613]
10. March S, Spence SH, Donovan CL. The efficacy of an internet-based cognitive-behavioral therapy intervention for child anxiety disorders. J Pediatr Psychol 2009 Jun;34(5):474-487 [FREE Full text] [doi: 10.1093/jpepsy/jsn099] [Medline: 18794187]
11. Spence SH, Donovan CL, March S, Gamble A, Anderson RE, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
12. Spence SH, Holmes JM, March S, Lipp OV. The feasibility and outcome of clinic plus internet delivery of cognitive-behavior therapy for childhood anxiety. J Consult Clin Psychol 2006 Jun;74(3):614-621. [doi: 10.1037/0022-006X.74.3.614] [Medline: 16822117]
13. Reyes-Portillo JA, Mufson L, Greenhill LL, Gould MS, Fisher PW, Tarlow N, et al. Web-based interventions for youth internalizing problems: a systematic review. J Am Acad Child Adolesc Psychiatry 2014 Dec;53(12):1254-1270.e5. [doi: 10.1016/j.jaac.2014.09.005] [Medline: 25457924]
14. 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 Interventions 2018 Jun;12:1-10. [doi: 10.1016/j.invent.2018.02.002]
15. Jolstedt M, Ljótsson B, Fredlander S, Tedgård T, Hallberg A, Ekeljung A, et al. Implementation of internet-delivered CBT for children with anxiety disorders in a rural area: A feasibility trial. Internet Interventions 2018 Jun;12:121-129 [FREE Full text] [doi: 10.1016/j.invent.2017.11.003]

16. Spence SH, Donovan CL, March S, Gamble A, Anderson R, Prosser S, et al. Online CBT in the treatment of child and adolescent anxiety disorders: issues in the development of BRAVE-ONLINE and two case illustrations. Behav Cognit Psychother 2008 Jul 01;36(04):411-430. [doi: 10.1017/s135246580800444x]
17. Australian Bureau of Statistics. Australian Standard Geographical Classification (ASGC). Canberra: Australian Bureau of Statistics 1216 2011.
18. Spence SH, Sawyer MG, Sheffield J, Patton G, Bond L, Graetz B, et al. Does the absence of a supportive family environment influence the outcome of a universal intervention for the prevention of depression? Int J Environ Res Public Health 2014 May 13;11(5):5113-5132 [FREE Full text] [doi: 10.3390/ijerph110505113] [Medline: 24828082]
19. Spence SH. A measure of anxiety symptoms among children. Behav Res Ther 1998 May;36(5):545-566. [doi: 10.1016/S0005-7967(98)00034-5]
20. Wolpert M, Görzig A, Deighton J, Fugard AJ, Newman R, Ford T. Comparison of indices of clinically meaningful change in child and adolescent mental health services: difference scores, reliable change, crossing clinical thresholds and 'added value' - an exploration using parent rated scores on the SDQ. Child Adolesc Ment Health 2014 Nov 24;20(2):94-101. [doi: 10.1111/camh.12080]
21. Rickwood DJ, Mazzer KR, Telford NR, Parker AG, Tanti CJ, McGorry PD. Changes in psychological distress and psychosocial functioning in young people visiting headspace centres for mental health problems. Med J Aust 2015 Jun 01;202(10):537-542. [Medline: 26021366]
22. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009 Mar;14(1):43-53 [FREE Full text] [doi: 10.1037/a0014699] [Medline: 19271847]
23. Muthén LK, Muthén BO. Mplus User's Guide. 8th edition. Los Angeles, CA: Muthén & Muthén; 2017.
24. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [Medline: 2002127]
25. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 02;2:97-105 [FREE Full text] [Medline: 19920949]
26. Chu BC, Skriner LC, Zandberg LJ. Shape of change in cognitive behavioral therapy for youth anxiety: symptom trajectory and predictors of change. J Consult Clin Psychol 2013 Aug;81(4):573-587. [doi: 10.1037/a0033390] [Medline: 23796317]
27. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 2015 Feb 06;350:h391. [doi: 10.1136/bmj.h391]

Abbreviations

ANOVA: analysis of variance
CAS-8: Children's Anxiety Scale, 8-item
CBT: cognitive behavioral therapy
iCBT: internet-based cognitive behavioral therapy
LGCM: latent growth curve modeling
RCI: Reliable Change Index
RCT: randomized controlled trials
UQ: University of Queensland

Edited by G Eysenbach; submitted 19.10.17; peer-reviewed by K Stasiak, E Law, L Donkin; comments to author 17.12.17; revised version received 05.04.18; accepted 10.05.18; published 04.07.18.

Please cite as:
March S, Spence SH, Donovan CL, Kenardy JA
Large-Scale Dissemination of Internet-Based Cognitive Behavioral Therapy for Youth Anxiety: Feasibility and Acceptability Study
J Med Internet Res 2018;20(7):e234
URL: http://www.jmir.org/2018/7/e234/
doi:10.2196/jmir.9211
PMID:29973338

Original Paper

Determinants of Successful eHealth Coaching for Consumer Lifestyle Changes: Qualitative Interview Study Among Health Care Professionals

Carl Joakim Brandt[1,2], MD; Gabrielle Isidora Søgaard[1], MPH; Jane Clemensen[2,3], RN, PhD; Jens Søndergaard[1], PhD; Jesper Bo Nielsen[1], PhD

[1]Research Unit of General Practice, Department of Public Health, University of Southern Denmark, Odense C, Denmark

[2]Centre for Innovative Medical Technology, University of Southern Denmark, Odense, Denmark

[3]Hans Christian Andersen's Children's Hospital, Odense University Hospital, Odense, Denmark

Corresponding Author:
Carl Joakim Brandt, MD
Research Unit of General Practice
Department of Public Health
University of Southern Denmark
JB Winsløwsvej 9A
Odense C, 5000
Denmark
Phone: 45 20141566
Email: cbrandt@health.sdu.dk

Abstract

Background: Success with lifestyle change, such as weight loss, tobacco cessation, and increased activity level, using electronic health (eHealth) has been demonstrated in numerous studies short term. However, evidence on how to maintain the effect long-term has not been fully explored, even though there is a pressing need for long-term solutions. Recent studies indicate that weight loss can be achieved and maintained over 12 and 20 months in a primary care setting using a collaborative eHealth tool. The effect of collaborative eHealth in promoting lifestyle changes depends on competent and skilled dieticians, nurses, physiotherapists, and occupational therapists acting as eHealth coaches. How such health care professionals perceive delivering asynchronous eHealth coaching and which determinants they find to be essential to achieving successful long-term lifestyle coaching have only been briefly explored and deserve further exploration.

Objective: The aim of this study is to analyze how health care professionals perceive eHealth coaching and to explore what influences successful long-term lifestyle change for patients undergoing hybrid eHealth coaching using a collaborative eHealth tool.

Methods: A total of 10 health care professionals were recruited by purposive sampling. They were all women aged 36 to 65 years of age with a mean age of 48 years of age. A total of 8/10 (80%) had more than 15 years of experience in their field, and all had more than six months of experience providing eHealth lifestyle coaching using a combination of face-to-face meetings and asynchronous eHealth coaching. They worked in 5 municipalities in the Region of Southern Denmark. We performed individual, qualitative, semistructured, in-depth interviews in their workplace about their experiences with health coaching about lifestyle change, both for their patients and for themselves, and mainly how they perceived using a collaborative eHealth solution as a part of their work.

Results: The health care professionals all found establishing and maintaining an empathic relationship essential and that asynchronous eHealth lifestyle coaching challenged this compared to face-to-face coaching. The primary reason was that unlike typical in-person encounters in health care, they did not receive immediate feedback from the patients. We identified four central themes relevant to the health care professionals in their asynchronous eHealth coaching: (1) establishing an empathic relationship, (2) reflection in asynchronous eHealth coaching, (3) identifying realistic goals based on personal barriers, and (4) staying connected in asynchronous coaching.

Conclusions: Establishing and maintaining an empathic relationship is probably the most crucial factor for successful subsequent eHealth coaching. It was of paramount importance to get to know the patient first, and the asynchronous interaction aspect presented challenges because of the delay in response times (both ways). It also presented opportunities for reflection before

answering. The health care professionals found they had to provide both relational communication and goal-oriented coaching when using eHealth solutions. Going forward, the quality of the health care professional–patient interaction will need attention if patients are to benefit from collaborative eHealth coaching fully.

(J Med Internet Res 2018;20(7):e237) doi:10.2196/jmir.9791

KEYWORDS

behavior change; eHealth coaching; empathy; lifestyle; healthy lifestyle; mHealth; mobile health units; primary health care; primary care; public health; relationship; telemedicine

Introduction

Background

Successful electronic health (eHealth) lifestyle coaching to increase exercise, improve diet, and reduce tobacco and alcohol use has been demonstrated in numerous studies [1].However, maintaining the effect over extended periods of time has had more variable results [2]. New studies have demonstrated remission from a diabetic state for almost half of a patient population solely by increased activity, diet, and weight loss in both primary and secondary care settings [3]. A recent study showed that 96% of a representative sample of 1004 Danes between 40-60 years of age preferred lifestyle change to medication [4] even though few general practitioners recognize this [5].

Many studies show that empathy by the health care professional (HCP) providing the lifestyle coaching is of paramount importance for in-person coaching [6,7]. Previously, we reported on a collaborative eHealth solution that resulted in long-term behavioral change where weight loss of 7.0 kg over 20 months was achieved using eHealth coaching in a general practice setting [8]. The same findings were observed in a municipality setting with diabetic men, where patients stated that an initial in-person meeting with the dietician seemed critical for their future Web-based interaction [9]. Other studies suggest that HCPs enjoy in-person meetings more than eHealth coaching [10]. Despite the success of these smaller-scale studies, there is a need to clarify various aspects of eHealth coaching and factors influencing successful long-term lifestyle change [11]. Use of eHealth is viewed positively by general practitioners (GPs), who use motivational interviewing in their practices and eHealth for their health [5].

The importance of the HCPs' support of patients with lifestyle challenges and how the HCPs perceive the use of eHealth has not yet been explored [12]. Hence, we aimed to identify factors essential to HCPs assisting patients undergoing lifestyle changes using eHealth. Of particular focus was how the HCPs viewed their eHealth coaching, what motivated them, and which factors in their eHealth coaching were most important for supporting their patients and guide them through the challenges faced on the way towards a healthier lifestyle.

Methods

Context

Denmark and the Danish health care sector have 3 political and administrative levels: the national state, 5 geographically defined regions, and 98 municipalities. Municipalities have on average approximately 57,000 inhabitants. They are local administrative bodies and deliver public health care, disease prevention, and rehabilitation at the local level, outside of hospitals [13].

Design

This qualitative study was based on in-depth and semistructured individual interviews with 10 Danish HCPs who provide eHealth coaching in health care centers in 5 municipalities in the Region of Southern Denmark. HCPs in a municipality health care center can have different health care education backgrounds including dieticians, physiotherapists, nurses, and occupational therapists.

Sampling

Sampling was conducted among 12 female HCPs providing eHealth coaching, who had coached more than 30 patients for more than 3 months, and individuals were recruited by email or phone. In total, 11 HCPs were invited, although 1 declined to participate due to a job change. Saturation was met after 7 interviews, but the remaining 3 interviews were conducted to confirm that no new themes or subthemes emerged [14]. The HCPs interviewed were all female, between 36-65 years of age, with a mean age of 48 years of age. A total of 8/10 (80%) had more than 15 years of experience in their field. All had experience providing hybrid eHealth lifestyle coaching using a combination of synchronous face-to-face meetings and asynchronous eHealth coaching through a collaborative eHealth tool. There were 10 female HCPs, including 5/10 (50%) clinical dieticians, 2/10 (20%) physiotherapists, 1/10 (10%) nurse, 1/10 (10%) occupational therapist, and 1/10 (10%) nurse assistant. Half 5/10 (50%) had taken specific postgraduate coaching courses in motivational interviewing, and 2/10 (20%) had other pedagogic educations. They had between 0.5-31 years of coaching experience. A total of 9/10 (90%) had other tasks, such as coaching or teaching patients in traditional face-to-face coaching or group sessions. They spent 4-16 hours per week on asynchronous eHealth coaching, and interacting with 20-140 current patients through the collaborative eHealth tool.

Interview Procedure

An explorative approach was followed in order to explore the HCPs´ subjective experiences and interpretations of working with eHealth coaching, focusing on motivational factors for a successful long-term lifestyle change. Semistructured interviews were conducted with the participating HCPs, following a basic, loose interview guide with overall fields of interest and probing questions that permitted in-depth exploration of the HCPs´ views and perceptions (see Table 1). The question guidelines helped the researcher (CJB) to follow an iterative approach with room for exploration of emerging themes and perspectives that could

be further explored in interviews with subsequent participants [15].

The interviews were carried out in the HCPs' offices from May to June 2017 and took 45-75 minutes each. All interviews were performed by CJB, who has worked as a GP for more than ten years and with different eHealth solutions for more than fifteen years.

Ethical Considerations

The ethics committee for the Region of Southern Denmark considered that the protocol could be approved and determined that the Medical Research Involving Human Subjects Act does not apply to this study [16]. All participants were informed of CJB's role as a GP, and shareholder of Liva Health Care A/S that delivered part of the software. It was emphasized that CJB would interview them as a researcher.

Before an interview was initiated, CJB briefly explained the nature of the research, answered any questions regarding the study, and described the study in layman's terms. The participants were informed of their rights, and CJB explained that the interview data would be anonymized. Both the participant and CJB signed informed consent documents. Emails and phone numbers were obtained from the municipalities before the study commenced. Researcher CJB invited the participants, made arrangements for the interviews and handled all phone calls and email correspondence regarding this matter.

Table 1. Interview guide for semistructured interviews with health care professionals.

Fields of interest	Probing questions
Experience with health coaching involving patients with lifestyle challenges in the municipality health center	Please tell me about good and bad experiences you have had with health coaching. Why do you think it played out that way?
Their own lifestyle experiences	• Have you ever taken the initiative to improve or change your lifestyle? • Do you use experiences from your own life in your coaching? Why or why not?

Fields of interest	Probing questions
Experience with eHealth and digital coaching in relation to their own and patients' health challenges	• How much experience do you have with communicating with patients using digital tools? • What works well and what does not work well in digital coaching?

Intervention

HCPs conducted eHealth coaching using the collaborative eHealth solution LIVA [17] in a hybrid manner, combining face-to-face meetings with eHealth coaching. LIVA is a refinement of the former eHealth solutions Slankedoktor.dk [8] and mydietician.org.uk [9], which were used and described in detail in 2 previously reported studies [8,9]. The 5 participating municipalities have offered this eHealth tool to patients for 6-12 months and have each included 100-400 patients. Patients using the eHealth solution report on individual goals in real-time including activity, diet, sleep, pain, and compliance with personal goals or other goals agreed on with the HCP, via iOS, Android or web. HCPs used a Web-based "backend" interface that served as a control panel, cockpit, and library. eHealth coaching is conducted asynchronously via short message service text messaging, or video messaging weekly, biweekly, monthly, or in a way the HCP decided was most appropriate to meet the patients' needs. In Multimedia Appendix 1, the eHealth solution LIVA is presented with detail inspired by the Template for Intervention Description and the Replication checklist [18], with information on the specific behavioral change techniques from the Coventry, Aberdeen and London-Refined taxonomy [19].

Analyses

The 10 interviews were digitally recorded and transcribed verbatim. Analyses were performed by the researchers (CJB, GIS, JC, JBN, and JS) using thematic analysis. An explorative approach of systematic text condensation was applied [20,21]. The analysis process began with all researchers reading through the transcripts. They gained their impressions of what they viewed as relevant and exciting themes and then met several times to discuss their different views and agree upon a "codebook" of categorized ideas and topics within specific themes and subthemes relevant for the set objectives. The researchers CJB and GIS then started the a priori coding of each transcript in the software program NVivo 11 Pro for Windows [22]. This was performed using a node structure that reflected identified themes and subthemes and allowed for expansion and reduction along the way. To make sure that the researchers coded, sorted and categorized the data in the same way—by identifying similar expressions, patterns, and sequences in the transcripts—the coding comparison function in NVivo 11 Pro was used on the first 3 interviews, and then coding was aligned

where necessary. The data from each of the identified themes were then condensed and summarized into generalized descriptions and concepts. In the analysis process, the researchers related the extracted information to the full transcripts to make sure they preserved the original context. The identified themes were compared between the different researchers several times throughout the process. In the end, these descriptive themes were put into analytical themes according to the thematic synthesis approach [23]. Finally, the quotes that best illustrated each theme and its related subthemes were selected and translated from Danish to English. The researchers CJB and GIS initiated the translation process by comparing their translations, agreeing on wording and meaning in the sentences, and then comparing them a second time to the Danish quotes. The remaining authors then reviewed all quotes in Danish and English, and changes were made if all parties agreed. In the text, interview quotes are followed by a unique participant identifier, ranging from Health Care Professional 1 to Health Care Professional 10.

Results

Themes and Subthemes

We identified 4 central themes with many subthemes concerning the HCPs' perceptions of conducting eHealth coaching (see Textbox 1 for an overview of these themes and their related subthemes):

- Establishing an empathic relationship
- Reflection in asynchronous eHealth coaching
- Identifying realistic goals based on personal barriers
- Staying connected in asynchronous coaching

Textbox 1. Themes and subthemes for using a collaborative electronic health (eHealth) tool in combination with face-to-face consultations for health care professionals.

Theme 1: Establishing an empathic relationship

- Combining synchronous face-to-face coaching with asynchronous eHealth coaching
- Use the health care professional's own story of lifestyle change
- Appreciating the communication in asynchronous eHealth coaching
- Health care professional's motivation

Theme 2: Reflection in asynchronous eHealth coaching

- Health care professional reflection
- Patient reflection
- Explore individual motivation

Theme 3: Identifying realistic goals based on personal barriers

- Recognize harmful patterns
- Operational goal setting
- Appreciate small steps

Theme 4: Staying connected in asynchronous coaching

- Personal comments
- Reading the patient
- Feedback stimulated by open questions

Establishing an Empathic Relationship

All HCPs found it challenging to provide proper eHealth coaching because it was not possible to get face-to-face feedback. Their typical tools to elicit feedback, such as mirroring body language or prompting patients to continue a line of thought by repeating the last word in a sentence, were not applicable since they were separated from their patients in time and space.

Combining Synchronous Face-to-face Coaching with Asynchronous eHealth Coaching

All HCPs found it essential to build an empathic relationship with room for reflection in face-to-face meetings before asynchronous eHealth coaching.

> *In this relationship we have built up, (by meeting face-to-face initially), they will tell you more personal things—at least that is what I experience—more than they did earlier on with Slankedoktor (digital-coaching only). I have coached one that has admitted excess eating, one that has told me that her daughter is at a crisis center, and it was very natural for them to share this.* [Health Care Professional 10]

Use the Health Care Professional's Own Story About Lifestyle Change

To get to know the patient better and connect with the patients, HCPs found it beneficial to tell their own stories about lifestyle change. A total of 9/10 (90%) HCPs found it essential to show the patients that they knew that lifestyle changes required hard

work. There were 7/10 (70%) HCPs that used their own experiences when they explained to patients what was needed to achieve a specific outcome. They often also made an effort to explain that despite looking healthy and fit, they also experienced challenges on a daily basis in maintaining their good health.

> *It is important for me to tell them that I am no bikini model, and I have been 25 kg heavier than I am now, so this is to say I know the kind of problems that matter on a daily basis.* [Health Care Professional 4]

Appreciating Communication in Asynchronous eHealth Coaching

When conducting eHealth coaching after the initial face-to-face meeting, HCPs found it essential to send messages with positive expectations, not only for measurable outcomes but also for communication itself, which was viewed as critical for patients to stay connected.

> *The patients tell me things like "I am always happy when I read what you have sent to me…"; "I look forward to seeing what you have written…" and one said"I am always so excited to see if you have found something I have done right.*" [Health Care Professional 7]

> *When I do digital coaching…I recognize that it has been difficult for the patient, or praise when I can see they are doing good. It can be a few words I send off or a video greeting. I really like video, because they say to me "I can feel you, it is like you are sitting on my shoulder cheering!"* [Health Care Professional 8]

Half of the HCPs (5/10, 50%) experienced nonjudgmental communication in "neutral waters". For example, greetings for the holiday seasons, resulted in many more responses in comparison to when they asked for performance data or sent out standard messages with health educational content.

> *Some of the patients I had not heard from, but then I wrote that I had to take a leave the next two weeks because I had broken my arm, then it was almost everyone who commented and wished me good health.* [Health Care Professional 1]

The Health Care Professional's Motivation

A total of 8/10 (80%) HCPs explained that they were motivated by meeting with another person, establishing a relationship, and getting closer to an "understanding" and "feeling" of the person in front of them. Even though many patients today are accustomed to digital communication, all HCPs found that an initial face-to-face meeting before initiating digital coaching was necessary to establish a strong and compassionate relationship.

> *I think that I was the factor that made the difference, since he (the patient) knew that I was the person who was coaching him. He had met me in person and it made a difference that it was not just another app he could use for entering his data. Here, he actually got concrete answers to his questions.* [Health Care Professional 9]

The coaching could only work if the patient communicated, and losing feedback was therefore seen as a significant challenge for all the HCPs.

> *Well, it motivates me when I get some kind of feedback from the patients. Then I think it is fun and nice to spend time on it. Those who do not give very much can be less motivating, I think.* [Health Care Professional 10]

Reflection in Asynchronous eHealth Coaching

The HCPs found that they could deliver advice with the use of very little time working asynchronously. A total of 9/10 (90%) HCPs said that they used only 5-10 minutes for each digital coaching session. In comparison, face-to-face coaching tends to be very time consuming for both the patient and the HCP (ie, 30-60 minutes). The lack of direct patient interaction in asynchronously coaching challenges the coach's abilities to see the patient's reaction to advice or questions directly. However, this opens for reflection for both the HCP and the patient. Individual motivation also needs to be explored in manners other than known from traditional motivational interviewing.

Health Care Professional Reflection

Due to the time difference between when patients enter data and when advice is given, the HCPs could think, reflect, and adjust their advice before sending it to the patient. The HCPs did not have to answer immediately when they saw data from the patient. Instead they could go for a walk or answer another question before they returned to give personalized advice regarding what they had observed.

> *So you can stop (your digital consultation) and reflect: "What is it she really needs?" …Then you can come back later and finish your consultation.* [Health Care Professional 7]

Patient Reflection

In the same manner, when advice was given, patients had time before responding, which could be seen as a chance to think, reflect, comment, or enter other data. There was 1/10 (10%) HCP who explained how she saw this as an advantage for the patient when difficult topics need to be dealt with:

> *One patient once told me: "You can write it in small pieces if it really hurts (eg, difficult to talk about), as opposed to when you meet at the doctor's surgery, at the dietician or at the psychologist you need to finish, you must say everything in the consultation right away.You cannot take a break, think about it and reply…"* [Health Care Professional 7]

Explore Individual Motivation

All HCPs found it essential to find out what motivated the patient. To accomplish this, they found it very important to give the patient space to reflect and initiated the coaching by providing the patient time for goal setting and reflection. Learning could then come from the lived life.

> *This man had diabetes, and he knew all about it, but he lacked ownership, and he did not understand how to cope with it. So, after he began here in the*

> *municipality center and we found out what help he needed, he began exercising and measuring his blood sugar. So now he has lost 20 kg, and he sees how exercise and healthy eating affect his blood sugar. He is really motivated when he sees the immediate effect and it is thought-provoking that, actually, he has never really understood the effects of carbs on the blood sugar (until now).* [Health Care Professional 5]

Identifying Realistic Goals Based on Personal Barriers

During the digital coaching sessions, 9/10 (90%) HCPs found it essential to getting to know the patient better to understand if they had destructive patterns and to identify realistic goals. This helped them to recognize patients' progress even though the patients did not see it themselves.

Recognize Harmful Patterns

The HCPs were often occupied with helping patients break free from harmful patterns and actions.

> *So, what can be the reason a person chooses to say that: "I cannot do it because of this and that." That is to say, what is the reason he only sees barriers, and is it a pattern he has had throughout life?* [Health Care Professional 1]

Operational Goal Setting

The collaborative eHealth tool supported specific goals set out by the patient. Helping patients to be concrete and operational in their goal setting was mentioned by 5/10 (50%) HCPs as a challenge. As an example, moving from the generic *"I want to live healthy"* to the specific *"I want to eat breakfast"* was of vital importance when patients monitored their daily performance; operational goal-setting was crucial to turning goals into measurable outcomes.

> *Sometimes they are just not precise enough. Some of them might want to "eat healthy", but what is it exactly they want to change? They need to be more concrete and specific about their challenges. Is it snacking in between meals that needs to be changed? Or what is it?* [Health Care Professional 5]

Appreciate Small Steps

In coaching sessions, 6/10 (60%) HCPs found it was important to recognize small signs of progress that might not be noted by the patient.

> *The patient could say: "I have not done anything since we last spoke". When you then look closer and see that they have done something, but just not reached the goals they had expected… So, you move focus to their successes.* [Health Care Professional 7]

Staying Connected in Asynchronous Coaching

All HCPs found it quite challenging when the patient did not respond to the eHealth coaching: if they took a very long time to reply, did not register their activities on a regular basis, or did not respond to the advice given in the last coaching session. A total of 9/10 (90%) HCPs explained that the lack of feedback often paused the process and made the HCP wonder what was going on—a situation that was new to them and indicated a need to approach things differently from what they had been used to in face-to-face coaching sessions.

> *So yes, using a collaborative eHealth tool really requires patience, because it takes a long time to get the answers. So, it has also been a process that has stretched over a long period of time, where I have asked her (the patient) a question, and I have added some reflective notes to it. And then I have waited for her answer before I could go on with the process. So, it is a different form, but I actually think that it has worked, yes! But you need to learn to accept, especially in the beginning, that it takes a long time, and that it is okay.* [Health Care Professional 3]

The HCP's had developed many strategies to stay connected through personal comments, reading the patient, and using open questions.

Personal Comments

The eHealth solution provided the opportunity for the HCP to reuse "standard advice". The HCPs explained that more than 50% of the content provided as either written or video advice that was reused. There were 6/10 (60%) HCPs explained that they made an effort to craft a unique, personalized, nonjudgmental frame around the necessary standard advice.

> *the specific advice is about 80% reuse, but I do make some small adaptations.* [Health Care Professional 2]

> *The tone I answer in will be unique and tailored to the individual—articles and recommendations will be reused of course—but the frame around it will always be unique. And then there will be prefabricated elements which are the same for everybody because it cannot be said in any other way.* [Health Care Professional 7]

Reading the Patient

The eHealth solution also allowed both the HCP and the patient to go back to an earlier question or answer to clarify what had been communicated. The HCPs highlighted this as positive for the interaction and very useful in situations where the HCP was uncertain about whether the patient shared the HCP's view of the content of the communication.

> *I start looking for their registrations to see if there is something positive to comment on. Then I often start there…when I have said something or done something in the last communication we have had, then in the next message I send out to the person I kind of remind myself that I have to ask them whether they found it useful or not, just to give them the chance to say: "Well… I think it was a bit far out!"* [Health Care Professional 6]

Feedback Stimulated by Open Questions

Feedback had to be stimulated in different ways than what HCPs were used to in face-to-face interactions. Most HCPs tried to

encourage more frequent feedback by sending very open but positive questions and remarks.

> *I asked the open question: "What do you eat and when do you eat?" and then I let her tell me herself. Then I asked her: "I can see you mention something you call junk and unhealthy stuff." Then she replies: "I do not really eat much of that, but I eat large portions. My stomach has been accustomed to that.*" [Health Care Professional 1]

> *I always praise them for the work: "well done" and make an effort to write my reply so that it mirrors the themes they have mentioned as important to them.* [Health Care Professional 10]

Discussion

Establishing and maintaining an empathic relationship with the patients was the single most crucial factor for the HCPs when they performed asynchronous eHealth coaching. This is in line with findings suggesting that lack of an empathic relationship with the patient can be toxic when providing motivational coaching [6].

Establishing an Empathic Relationship

Empathy is an independent contributor to the benefit of behavioral interventions [24]. However, empathy is difficult to maintain in eHealth due to the need for mutual confirmation that happens through signs and signals when coaches interact with someone from their own "tribe" [25]. This tribal verbal and nonverbal language is fundamentally challenged by communicating digitally. The Internet provides access to an information overload that can be difficult to interpret for patients with low health literacy [26]. Our study suggests that if a trustworthy relationship is established and maintained, HCPs using hybrid eHealth coaching methods could be very useful for patients with low health literacy.

A systematic review of previous systematic reviews of studies using Web-based weight loss interventions revealed conflicting results for effects when comparing Web-based interventions with hybrid interventions [2]. Earlier studies on the consultation process revealed that health professionals only have access to a patient's reflections on difficult, personal and relevant subjects about their health if the HCP manages to establish an empathic relationship [27]. Using hybrid, complex interventions, meeting patients both in-person (ie, synchronous) to strengthen relations and through asynchronous eHealth might improve health care through more effective and efficient interpersonal communication, even though long-term studies beyond 24 months are still missing [1,9]. A pilot randomized controlled trial study on feasibility and acceptability of a prior, Web-based version of the asynchronous eHealth tool used in this study for men with type 2 diabetes revealed that eHealth can facilitate relevant in-time feedback and profound reflections from the patients. Moreover, the effect of the intervention seems better when an empathic relationship between the patient and the HCP is established before the Web-based intervention is initiated [9]. HCPs used stories about their own health challenges to find common ground with the patients. For decades, health professionals have been warned against using their own personal health experiences in patient treatment. Moving into the 21st century, this notion might be challenged by successes with collaborative hybrid eHealth solutions, where empathy between patients and HCPs is a pivotal factor in securing long-term success [24]. Empathy may prove difficult to establish without HCPs using personal stories. We found that HCPs, who repeatedly appreciated communication through positive reinforcement of the asynchronous communication and not only the measurements registered seemed most successful in engaging patients and maintaining an empathic relation. Appreciating patient communication through text and video actively might be the HCPs way to express "reflective listening" digitally, known in the behavioral change theory [6]. The success of traditional motivational coaching depends on the empathy of the HCP [6]. Our study shows that empathy by the HCP can be challenged in asynchronous eHealth coaching. It, therefore, seems to be of paramount importance also to focus on what motivates the HCP and how their self-empathy can be nourished [25]. One of the critical determinants found in our study is the need for feedback by the patients to the HCP for the HCP to stay personal.

Reflection in Asynchronous eHealth Coaching

In the current study, the HCPs described how the asynchronous eHealth coaching provided both patients and HCPs with the chance to reflect during the time interval between questions and answers. The value of having time for reflection has not, to our knowledge, been evaluated in other eHealth studies. Most studies investigating asynchronous eHealth communication have examined how primary care providers can optimize their access to specialists [28]. An earlier systematic review looking at Web-based solutions compared asynchronous with synchronous eHealth consultations [29] Unfortunately, they only found a few studies of relevant quality, which meant that they could not make any conclusions about asynchronous versus synchronous solutions.

Identifying Realistic Goals Based on Personal Barriers

Barriers to lifestyle changes can be difficult to detect. We speculate that these barriers change over time, and what seems realistic in an initial in-person meeting in a municipality setting might look different when reality strikes at home. This is an issue that might have a more significant impact on the lives of patients of low socioeconomic status [30]. We found that patients often perceive goal-setting as a process of creating long-term, distant and broader goals, which can give them a feeling of defeat if they fail. An essential part of the HCP's job in eHealth is to assist patients in setting realistic, short-term, measurable subgoals to make sure distant goals do not decrease the patient's self-efficacy. Self-efficacy is known to be a critical success factor for a lifestyle change. Apart from this, emphasizing the "small victories" is a key strategy used by many of the eHealth coaches, which makes good sense since self-efficacy can be increased by increasing a person's feeling of control over the behavior. When these goals are appropriately structured, eHealth is unique in its ability to help individuals achieve measurable, realistic goals [31].

Staying Connected in Asynchronous Coaching

Most of the HCPs revealed how important it was to be personal in their communications. Earlier studies have shown a high adherence to hybrid asynchronous collaborative eHealth coaching tools, which might be due to patients taking responsibility and feeling in charge [8,9]. In this study, when the relationship was inactive due to the patient not interacting, the HCPs found many ways to re-establish communication, such as being personal, reading the patient and using open questions. Being a trusted person involves providing nonjudgmental emotional support through conversation, reflective listening, touch, and physical presence when helping patients and relatives through difficult times [32]. Translating this into eHealth coaching is challenging, but personal nonjudgmental coaching seems essential in order not to lose patient feedback. One important takeaway message from the HCPs is that no matter how trivial a message or a video might seem, in order to make a difference for the patient it should always be personalized if possible. Personalization can be done by using information found in the message dialogue history. The need for personalization in asynchronous coaching is in line with other studies examining the delivering of standardized information [33]. The possibility of sending relevant and timely individual videos both from HCPs and patients might also be of vital importance. Using asynchronous video messaging might be of special importance in breaking down some of the obstacles presented by low health literacy. The question of how this would affect motivation and change behavior is still not scientifically documented.

Strengths and Limitations of this Study

This is the first qualitative research study to analyze how HCPs, coaching via a hybrid, collaborative eHealth tool, perceive what is essential for successful lifestyle change among patients. The findings of this study are relevant and are expected to be of more general use in future research regarding the effectiveness and implementation challenges of collaborative eHealth solutions. However, in our study all HCPs were female, and even though saturation of central themes and subthemes was achieved in this group, more heterogeneity of eHealth coaches may lead to additional insights. Further dissemination of the eHealth tool used here along with other collaborative eHealth tools will also demand more research as they will not be applicable in all health care systems.

A limitation of this study is also the lack of methodological triangulation, as we only studied the perspective of the health care professionals and did not examine the patient perspective nor quantify the aspects revealed. For this reason, further studies using questionnaires and quantitative outcomes are suggested.

Conclusion

Successful eHealth coaching requires establishing and maintaining an empathic relationship. HCPs found it of paramount importance to get to know the patient first, preferably in an initial face-to-face meeting and to provide both relational communication and goal-oriented coaching when using eHealth solutions. The asynchronous interaction aspect presented challenges because of the delay in response times (ie, both ways), but it also presented opportunities for reflection before answering. The future quality of the HCP-patient interaction will need attention if patients are to fully benefit from behavior change techniques made possible by eHealth coaching. Our findings suggest that it will be of great value to the future development of collaborative eHealth interventions if the quality of the HCP-patient interaction is taken further into account. This includes focusing on educating the health professionals about their empathic role as eHealth coaches and by strengthening their ability to communicate with empathy via new digital tools. This study emphasizes that collaborative eHealth tools used in empathic patient care can constitute an effective way to deliver health care services compassionately in the future for patients needing to implement lifestyle changes.

Acknowledgments

The study has been partly funded by the Region of Southern Denmark, KEU project 07/10 and the Region of Southern Denmark.

Conflicts of Interest

The corresponding author CJB owns shares in Liva Health Care AS, the company that has developed the technical platform, LIVA, used in the study. The other authors declare that they have no personal financial interests related to the subject matters discussed in the manuscript.

Multimedia Appendix 1

The Template for Intervention Description and the Replication checklist for the eHealth solution LIVA.

[PDF File (Adobe PDF File), 40KB - jmir_v20i7e237_app1.pdf]

References

1. Afshin A, Babalola D, Mclean M, Yu Z, Ma W, Chen C, et al. Information Technology and Lifestyle: A Systematic Evaluation of Internet and Mobile Interventions for Improving Diet, Physical Activity, Obesity, Tobacco, and Alcohol Use. J Am Heart Assoc 2016 Aug 31;5(9):e003058 [FREE Full text] [doi: 10.1161/JAHA.115.003058] [Medline: 27581172]

2. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-Based Interventions for Weight Loss or Weight Loss Maintenance in Overweight and Obese People: A Systematic Review of Systematic Reviews. J Med Internet Res 2017 Jun 26;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
3. Lean ME, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet 2018 Dec 10;391(10120):541-551. [doi: 10.1016/S0140-6736(17)33102-1] [Medline: 29221645]
4. Jarbøl DE, Larsen PV, Gyrd-Hansen D, Søndergaard J, Brandt C, Leppin A, et al. Determinants of preferences for lifestyle changes versus medication and beliefs in ability to maintain lifestyle changes. A population-based survey. Prev Med Rep 2017 Jun;6:66-73 [FREE Full text] [doi: 10.1016/j.pmedr.2017.02.010] [Medline: 28271023]
5. Brandt CJ, Søgaard GI, Clemensen J, Sndergaard J, Nielsen JB. General Practitioners' Perspective on eHealth and Lifestyle Change: Qualitative Interview Study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e88 [FREE Full text] [doi: 10.2196/mhealth.8988] [Medline: 29666045]
6. Moyers TB, Miller WR. Is low therapist empathy toxic? Psychol Addict Behav 2013 Sep;27(3):878-884 [FREE Full text] [doi: 10.1037/a0030274] [Medline: 23025709]
7. Brandt CJ, Clemensen J, Nielsen JB, Søndergaard J. Drivers for successful long-term lifestyle change, the role of e-health: a qualitative interview study. BMJ Open 2018 Dec 12;8(3):e017466 [FREE Full text] [doi: 10.1136/bmjopen-2017-017466] [Medline: 29530904]
8. Brandt V, Brandt C, Glintborg D, Arendal C, Toubro S, Brandt K. Sustained weight loss during 20 months using a personalized interactive internet based dietician advice program in a general practice setting. International Journal on Advances in Life Sciences Internet. (1&2) pp 2011;3:23-28 [FREE Full text]
9. Haste A, Adamson AJ, McColl E, Araujo-Soares V, Bell R. Web-Based Weight Loss Intervention for Men With Type 2 Diabetes: Pilot Randomized Controlled Trial. JMIR Diabetes 2017 Jul 07;2(2):e14 [FREE Full text] [doi: 10.2196/diabetes.7430]
10. Smith E, Bradbury K, Scott L, Steele M, Little P, Yardley L. Providing online weight management in Primary Care: a mixed methods process evaluation of healthcare practitioners' experiences of using and supporting patients using POWeR. Implement Sci 2017 Dec 25;12(1):69 [FREE Full text] [doi: 10.1186/s13012-017-0596-6] [Medline: 28545538]
11. Collins C, Rochfort A. Promoting Self-Management and Patient Empowerment in Primary Care. In: Primary Care in Practice-Integration is Needed. -: InTechOpen; 2016:27-42.
12. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
13. Danish Ministery of health. eHealth in Denmark - eHealth as part of a coherent Danish health care system. Copenhagen: Danish Ministry of Health; 2012. eHealth in Denmark - eHealth as part of a coherent Danish health care system URL: http://www.sum.dk/~/media/Filer [accessed 2017-10-11] [WebCite Cache ID 6u8HrwsAe]
14. Mason M. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. 2010 2010-08-24 2010;11(3):art 8. [doi: 10.17169/fqs-11.3.1428]
15. Kvale S, Brinkmann S. Interviews: Learning the Craft of Qualitative Research Interviewing. 3rd ed. United States of America: SAGE; 2015.
16. Yip C, Han NR, Sng BL. Legal and ethical issues in research. Indian J Anaesth 2016 Sep;60(9):684-688 [FREE Full text] [doi: 10.4103/0019-5049.190627] [Medline: 27729698]
17. Koch C, From K, Bech R, Sode A, Brandt CJ, Bolting L, et al. Livahealthcare. URL: http://www.livahealthcare.com/index2.html [accessed 2017-10-10] [WebCite Cache ID 6u7DX0GUI]
18. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687 [FREE Full text] [Medline: 24609605]
19. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
20. Malterud K. Qualitative research: standards, challenges, and guidelines. Lancet 2001 Aug 11;358(9280):483-488. [doi: 10.1016/S0140-6736(01)05627-6] [Medline: 11513933]
21. Sandelowski M. Focus on research methods - whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [Medline: 10940958]
22. QSR International Pty Ltd. NVIVO 11 PRO for Windows. URL: http://www.qsrinternational.com/nvivo/nvivo-products/nvivo11-for-windows/nvivo-pro [accessed 2017-10-11] [WebCite Cache ID 6u82yFzq4]
23. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
24. Moyers TB, Houck J, Rice SL, Longabaugh R, Miller WR. Therapist empathy, combined behavioral intervention, and alcohol outcomes in the COMBINE research project. J Consult Clin Psychol 2016 Mar;84(3):221-229 [FREE Full text] [doi: 10.1037/ccp0000074] [Medline: 26795938]

25. Riess H. The Science of Empathy. Journal of Patient Experience 2017:2374373517699267.
26. Holmberg C, Berg C, Dahlgren J, Lissner L, Chaplin JE. Health literacy in a complex digital media landscape: Pediatric obesity patients' experiences with online weight, food, and health information. Health Informatics J 2018 Mar 01:1460458218759699. [doi: 10.1177/1460458218759699] [Medline: 29499615]
27. Zoffmann V, Harder I, Kirkevold M. A person-centered communication and reflection model: sharing decision-making in chronic care. Qual Health Res 2008 May;18(5):670-685. [doi: 10.1177/1049732307311008] [Medline: 18223158]
28. Tran CS, Liddy CE, Liu DM, Afkham A, Keely EJ. eConsults to endocrinologists improve access and change primary care provider behavior. Endocrine Practice 2016 Jun 13;22(10):1145-1150. [doi: 10.4158/EP161321.OR]
29. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van GL. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May;4(3):666-684 [FREE Full text] [Medline: 20513335]
30. Nielsen JB, Leppin A, Gyrd-Hansen DE, Jarbøl DE, Søndergaard J, Larsen PV. Barriers to lifestyle changes for prevention of cardiovascular disease - a survey among 40-60-year old Danes. BMC Cardiovasc Disord 2017 Sep 12;17(1):245 [FREE Full text] [doi: 10.1186/s12872-017-0677-0] [Medline: 28899356]
31. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
32. Michelson KN, Patel R, Haber-Barker N, Emanuel L, Frader J. End-of-life care decisions in the Pediatric Intensive Care Unit: roles professionals play. Pediatr Crit Care Med 2013 Jan;14(1):e34-e44 [FREE Full text] [doi: 10.1097/PCC.0b013e31826e7408] [Medline: 23249788]
33. Ellis R, Connor U, Marshall J. Development of patient-centric linguistically tailored psychoeducational messages to support nutrition and medication self-management in type 2 diabetes: a feasibility study. Patient preference and adherence 2014;8:1399.

Abbreviations

eHealth: electronic health
HCP: health care professional
GP: general practitioner

Edited by G Eysenbach; submitted 12.01.18; peer-reviewed by M Bardus, I Skärsäter, R Kornhaber, K Champion; comments to author 18.02.18; revised version received 27.04.18; accepted 15.05.18; published 05.07.18.

Please cite as:
Brandt CJ, Søgaard GI, Clemensen J, Søndergaard J, Nielsen JB
Determinants of Successful eHealth Coaching for Consumer Lifestyle Changes: Qualitative Interview Study Among Health Care Professionals
J Med Internet Res 2018;20(7):e237
URL: http://www.jmir.org/2018/7/e237/
doi:10.2196/jmir.9791
PMID:

Original Paper

Effectiveness of a Blended Care Self-Management Program for Caregivers of People With Early-Stage Dementia (Partner in Balance): Randomized Controlled Trial

Lizzy MM Boots[1], PhD; Marjolein E de Vugt[1], PhD; Gertrudis IJM Kempen[2], PhD; Frans RJ Verhey[2], PhD

[1]Alzheimer Center Limburg, Department of Psychiatry and Neuropsychology, School for Mental Health and Neurosciences, Maastricht University Medical Center, Maastricht, Netherlands

[2]Department of Health Services Research, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Lizzy MM Boots, PhD
Alzheimer Center Limburg, Department of Psychiatry and Neuropsychology
School for Mental Health and Neurosciences
Maastricht University Medical Center
PO Box 616
Maastricht, 6200 MD
Netherlands
Phone: 31 43 3881994
Fax: 31 43 3884092
Email: l.boots@maastrichtuniversity.nl

Abstract

Background: The benefits of electronic health support for dementia caregivers are increasingly recognized. Reaching caregivers of people with early-stage dementia could prevent high levels of burden and psychological problems in the later stages.

Objective: The current study evaluates the effectiveness of the blended care self-management program, Partner in Balance, compared to a control group.

Methods: A single-blind randomized controlled trial with 81 family caregivers of community-dwelling people with mild dementia was conducted. Participants were randomly assigned to either the 8-week, blended care self-management Partner in Balance program (N=41) or a waiting-list control group (N=40) receiving usual care (low-frequent counseling). The program combines face-to-face coaching with tailored Web-based modules. Data were collected at baseline and after 8 weeks in writing by an independent research assistant who was blinded to the treatment. The primary proximal outcome was self-efficacy (Caregiver Self-Efficacy Scale) and the primary distal outcome was symptoms of depression (Center for Epidemiological Studies Depression Scale). Secondary outcomes included mastery (Pearlin Mastery Scale), quality of life (Investigation Choice Experiments for the Preferences of Older People), and psychological complaints (Hospital Anxiety and Depression Scale-Anxiety and Perceived Stress Scale).

Results: A significant increase in favor of the intervention group was demonstrated for self-efficacy (care management, P=.002; service use P=.001), mastery (P=.001), and quality of life (P=.032). Effect sizes were medium for quality of life (d=0.58) and high for self-efficacy care management and service use (d=0.85 and d=0.93, respectively) and mastery (d=0.94). No significant differences between the groups were found on depressive symptoms, anxiety, and perceived stress.

Conclusions: This study evaluated the first blended-care intervention for caregivers of people with early-stage dementia and demonstrated a significant improvement in self-efficacy, mastery, and quality of life after receiving the Partner in Balance intervention, compared to a waiting-list control group receiving care as usual. Contrary to our expectations, the intervention did not decrease symptoms of depression, anxiety, or perceived stress. However, the levels of psychological complaints were relatively low in the study sample. Future studies including long-term follow up could clarify if an increase in self-efficacy results in a decrease or prevention of increased stress and depression. To conclude, the program can provide accessible preventative care to future generations of caregivers of people with early-stage dementia.

Trial Registration: Netherlands Trial Register NTR4748; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4748 (Archived by WebCite at http://www.webcitation.org/6vSb2t9Mg)

(J Med Internet Res 2018;20(7):e10017) doi:10.2196/10017

KEYWORDS

internet; caregivers; technology; therapeutics

Introduction

The majority of people with dementia are living at home and cared for by a family member, the informal caregiver. Informal care will be increasingly important as the number of people with dementia has been predicted rise to 65.7 million by 2030 and 115.4 million by 2050, together with a decrease of the working population [1].

However, informal caregiving has a downside. Caregivers of people with dementia are vulnerable due to the chronic stress they experience in the caregiving process [2], which may result in depression, anxiety, and other health problems [3]. Many caregiver support interventions have been developed to ameliorate negative caregiver consequences with promising results [4].

Early intervention and support for caregivers could prevent high levels of burden and psychological problems in the later stages of dementia [5,6]. However, early-stage interventions may not be effective, and even do more harm than good if they do not fit the personal situation of the caregiver. Negative and stigmatizing information can hamper acceptance, while enhancing the positive, intact experiences may be effective in increasing caregiver self-efficacy [7]. The Stress and Coping paradigm by Lazarus and Folkman [8] and the Social Learning theory by Bandura [9] propose that taking charge of the changes in one's life has a positive effect on self-efficacy and can therefore reduce caregiver stress and its negative impact on general wellbeing [10]. By increasing caregiver resilience through self-efficacy, an increase of psychological problems in a later stage may be prevented [9]. A self-management approach provides an excellent opportunity to actively involve caregivers and let them choose the themes and strategies that are best tailored to their needs. This suits the caring role transition in the early stages, which leans more towards a focus on positively managing life with dementia rather than managing the dementia itself [11].

With the growing gap between the number of people in need of support and available care professionals [12], electronic health (eHealth) interventions could serve as cost-effective alternatives for dementia caregiver support [13], with increased access and extended reach [14-17]. Blending face-to-face guidance with online support increases client-therapist connection and adherence [18,19]. Although eHealth interventions for caregivers have been developed and evaluated, so far most of them are aimed at dementia related problems in an advanced stage of the caregiver career [20,21] and their overall quality of evidence is low [22]. An iterative step-wise approach was employed to develop the blended care self-management internet-based Partner in Balance (PiB) program for caregivers of people with early-stage dementia. The current study evaluated if PiB is superior to a waiting-list control condition as evidenced by improved subjective self-confidence (self-efficacy and mastery), and lower levels of psychological complaints (symptoms of depression, anxiety, and stress) postintervention.

Methods

Overview

This randomized controlled trial was carried out between 2014 and 2016 in the Netherlands. The PiB program was compared to a waiting-list control group receiving usual care. Following the waiting-list period participants were offered the opportunity to follow the PiB program. The Medical Ethics Committee of the Maastricht University Medical Center+ (MUMC+) approved this study (#12-4-059) and the study was registered in the Dutch trial register (NTR4748). The study protocol and supporting SPIRIT checklist are available [23].

From September 2014 to December 2015, family caregivers of people with mild dementia of all subtypes (Clinical Dementia Rating, score 0.5-1) [24] were recruited from memory clinics (MUMC+, Elkerliek Hospital Helmond, Catharina Hospital Eindhoven) and ambulatory mental health clinics (Virenze-RIAGG Maastricht, MET ggz Roermond) in the south of the Netherlands. In addition, caregivers were informed about the trial via caregiver support services, and the website of the Dutch Alzheimer Association. Caregivers were included if they had access to the internet at home, had basic computer skills, and provided written informed consent. Potential participants with insufficient cognitive abilities to engage in the online self-management program, who were overburdened or with severe health problems as determined by study staff, or who cared for people with dementia caused by HIV, acquired brain impairment, Down syndrome, chorea associated with Huntington disease, or alcohol abuse were excluded from participation. Inclusion and exclusion was based on the clinical judgment of the referrer, based on their experience with the target group. Both spouses and other caregivers (eg, children) could be included, as long as they met the criteria above and were >18 years. Details on the recruitment procedure are described in the study protocol [23].

Randomization and Masking

Following the baseline assessment, participants were randomly assigned to either the PiB program or the waiting-list control group receiving usual care by the first author. Assignment was carried out using a computerized random-number generator for block randomization with variable sizes of 4, 6, and 8. An independent research assistant who was blinded to the allocation of the treatment conducted the postintervention assessments. It was not possible to blind the participants because of obvious differences between the interventions in content (PiB is a multicomponent intervention combining psycho-education, movie clips, assignments, and change plans and usual care often consists of psycho-education) and mode of delivery (PiB blends face-to-face contact with online modules and usual care often consists of face-to-face contact only).

Intervention and Control

Experimental Group: Partner in Balance

Detailed information about the program components and development is presented elsewhere [25]. In short, the blended care self-management program PiB consists of: (1) a face-to-face intake session with a personal coach to familiarize participants with the program, set goals, and select preferred module themes; (2) tailored online thematic modules, including psychoeducation, behavioral modeling, reflective assignments, change plans, and email feedback from the coach over 8 weeks; and (3) a face-to-face evaluation session with the coach evaluating previously set goals. All participants in the PiB group received these two face-to-face interactions with the personal coach. Furthermore, the participants can interact with other participants via a discussion forum. Module themes are acceptance, balance in activities, communication with family member and environment, coping with stress, focusing on the positive, insecurities and rumination, self-understanding, the changing family member, and social relations and support. Figure 1 shows a screenshot of the module themes in the program. The participants choose 4 modules and 2 weeks were allocated for each module. However, the participants were allowed to complete the modules at their own pace in accordance with the self-management approach [26]. The personal page and modules remained accessible for participants after the intervention period. The personal coaches were trained, experienced professionals (psychologists and psychiatric nurses) from one of the participating organizations. They attended a 2-hour training session in self-management techniques, goal setting, and online help and attended regular supervision meetings.

Figure 1. Screenshot of the module themes of the Partner in Balance program.

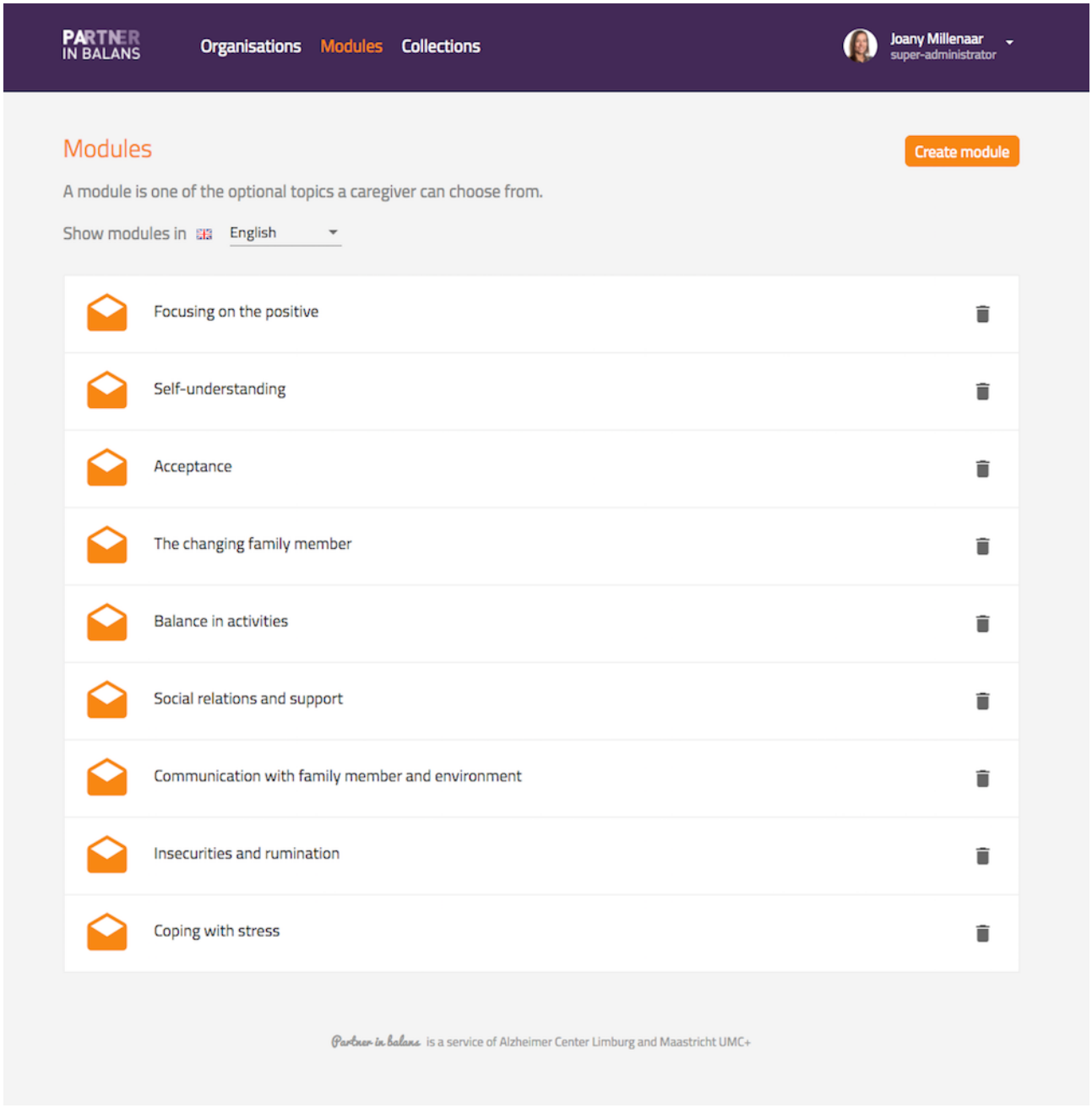
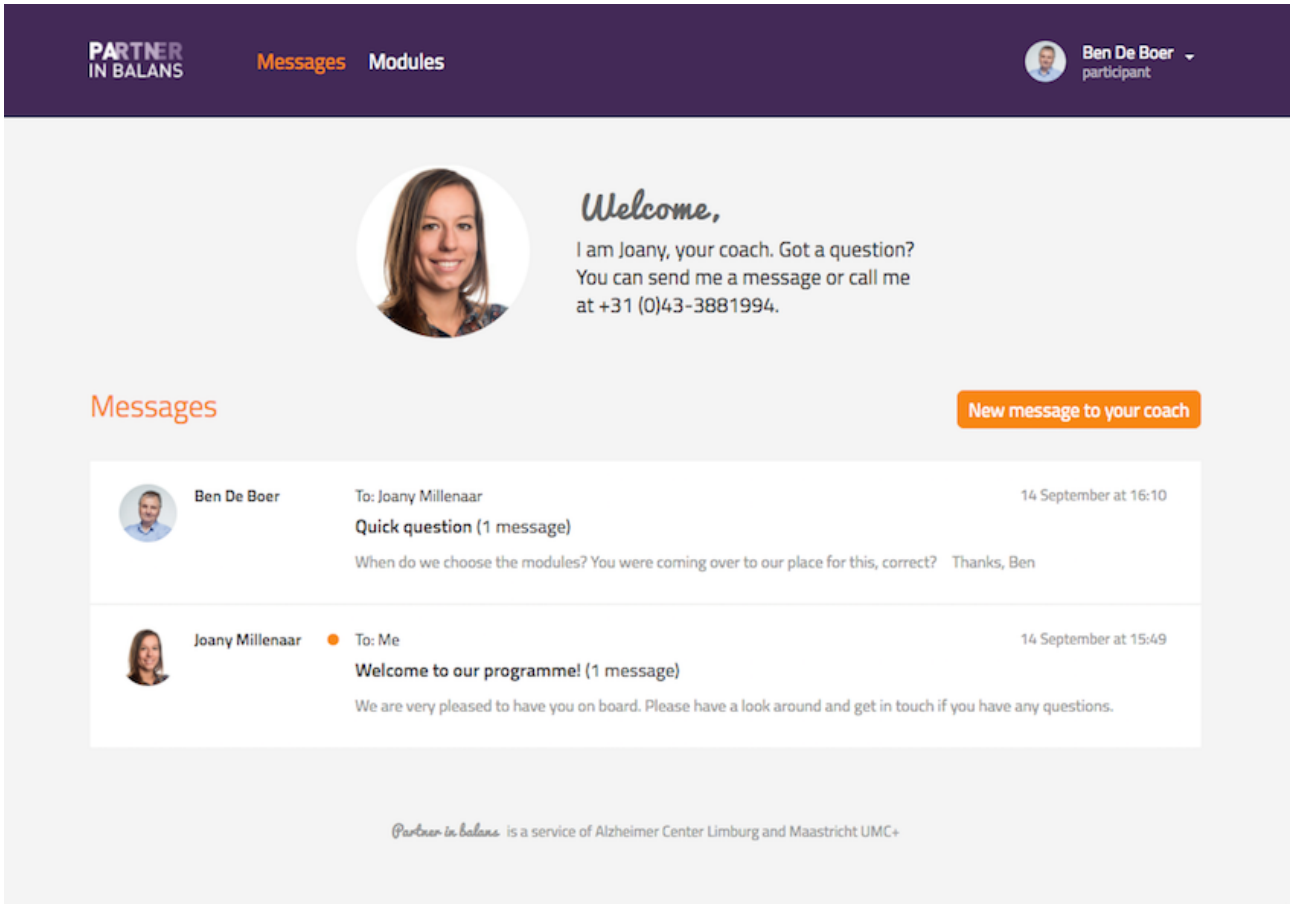


Figure 2. Screenshot of the Partner in Balance program's online messaging portal.



Their tasks were familiarizing participants with the online program, supporting them in module choice and goal setting, and providing feedback on the self-reflective assignments through the online messaging portal in the program (see Figure 2).

Control Group: Waiting List

The waiting-list group received usual care consisting of nonfrequent counseling during the 8 weeks. They received the same pretest and posttest attention from the research team as the experimental group. After they completed the posttest assessment, they were given the opportunity to follow PiB.

Procedures

For this study, self-reported data from the baseline visit (T_0) and after 8 weeks (T_1) were compared. These data were collected in writing by an independent research assistant who was blinded to the treatment, separately from the coach visits.

The primary proximal outcome was caregiver self-efficacy and primary distal outcome was depressive symptoms. Caregiver self-efficacy was measured with *The Caregiver Self-Efficacy Scale (CSES)* [27], measuring care management self-efficacy (4 items) and service use self-efficacy (5 items). Care management self-efficacy scores theoretically range from 4-40 and service use self-efficacy from 5-50. Higher scores on the CSES indicate higher levels of self-efficacy. The 20-item *Centre for Epidemiological Studies Depression Scale (CES-D)* [28] was used to measure depressive symptoms. Total scores range from 0-60; where higher scores indicate more symptoms. Secondary outcomes were mastery, psychological complaints (anxiety and perceived stress), and quality of life. Mastery was measured with the 7-item *Pearlin Mastery Scale (PMS)* [29]. The total score ranges from 7-35; where higher scores indicate higher levels of mastery. The 7-item *Hospital and Anxiety Depression Scale-Anxiety (HADS-A)* [30] rates symptoms of anxiety. Scores theoretically range from 0-21 with higher scores indicating more symptoms. Quality of life was measured on five attributes with the *Investigating Choice Experiments for the Preferences of Older People CAPability measure for Older people (ICECAP-O)* [31]. This index value indicates how good or bad the average person aged 65 or older considers a given state to be, for instance attributing to “attachment” (love and friendship) and “control” (independence). The value system for the 1024 (4^5=1024) possible states uses a best-worst scaling valuation method, providing a single summary score, anchored at zero (“no capability”) and 1.0 (“full capability”) [32].

Demographics were obtained (sex, age, relationship to care recipient, level of education, sharing household, and care intensity in years). The *Global Deterioration Scale* (GDS) [33] measured dementia severity with the caregiver as the informant. The possible modifying effects of the following variables were measures. Quality of the relationship was measured using 4 self-rating items of the University of Southern California Longitudinal Study of Three-Generation Families measures of positive affect [34]. The 12-item Emotional instability domain of the *NEO Five Factory Inventory (NEO-FFI)* [35], was used to identifying individuals who are prone to psychological distress, by assessing 6 traits: anxiety, angry hostility,

depression, self-consciousness, impulsiveness, and vulnerability. Scores ranged from 0-24; where higher scorers are likely to be sensitive, emotional, and more prone to experiencing feelings that are upsetting.

Sample Size

We aimed to enroll 80 participants (40 participants per group), based on previous online intervention studies in caregivers of people with dementia with the CSES as outcome measure, on the basis of repeated measures, within-between interaction with a mean effect size of 0.2 [36], assuming an alpha of .05, a power of 85%, and 25% loss to follow-up.

Data Analysis

Prior to the analysis, data were checked for missing values, outliers, and normality. Possible differences between the study groups' baseline characteristics were tested with *t* tests for continuous variables and chi-square tests for categorical variables. Nonparametric tests (eg, Mann-Whitney *U* Wilcoxon test) were used when necessary in case of nonnormality.

To examine the differences between outcomes for the intervention and the waiting-list control group during the intervention period, an analysis of covariance (ANCOVA) was conducted with outcome at post intervention as the dependent variable, intervention (PiB program, waiting-list control group) as the between-subjects variable and per outcome its baseline value, age, sex, emotional instability, quality of the relationship, educational level, and relationship to the care recipient as covariates. If significant, the intergroup effect size was calculated according to Cohen *d*. Effect sizes of 0.2 were considered small, 0.5 considered medium and 0.8 was considered high [37]. IBM SPSS statistics 22.0 for Macintosh was used and all tests of significance were two-tailed with alpha set at .05 and reported mean change.

User Involvement

As recommended by the Medical Research Council (MRC) Framework, a stepwise approach was adopted to explore potential user needs, followed by a pilot evaluation to test the feasibility of the intervention and the measurement tools prior to the effect evaluation. The iterative development and pilot evaluation of PiB as recommended by the MRC framework is described elsewhere [25]. The burden of the intervention was assessed in a process evaluation. Further, results were disseminated to study participants by means of a newsletter and PhD thesis.

Results

Participants

A total of 163 caregivers expressed an interest to participate. See Figure 3 for the study flowchart, the details of which are described elsewhere [38]. Table 1 lists the baseline data for the included caregivers (N=81).

Between-group comparisons revealed no significant differences in demographics and main outcome measures at baseline. Care recipients of the included caregivers were 73.9 years old (SD 8.2), diagnosed with Mild Cognitive Impairment (MCI; 12/81, 15%), Alzheimer's Disease (AD; 33/81, 41%), or other dementias (36/81, 44%). Dementia severity was rated as preclinical memory decline (55/81, 68%), mild dementia (24/81, 30%), or moderate dementia (2/81, 2%) on the GDS. At T_1, 13 caregivers were lost to follow-up. The completers did not differ from noncompleters at baseline in terms of age (t_{79}=0.19; *P*=.851), relationship to the care recipient (χ^2_1=1.39; *P*=.238), same household as care recipient (χ^2_1=0.82; *P*=.665), care intensity in years (*U*=377.5; *P*=.781), sex (χ^2_1=2.80; *P*=.094), education (χ^2_1=1.20; *P*=.550), self-efficacy service use (t_{79}=0.53; *P*=.599) care management (t_{79}=1.36; *P*=.177), depression (*U*=280.0; *P*=.266), stress (t_{79}=0.25, *P*=.806), anxiety (*U*=372.0; *P*=.497), mastery (t_{79}=–1.18; *P*=.253), and quality of life (*U*=775.0; *P*=.956).

Intervention Effects

The effects were compared between groups (intervention and waiting-list control) after 8 weeks. Table 2 shows the results of the ANCOVA at T_1 on self-efficacy (care management and service use), depression, mastery, perceived stress, anxiety, and quality of life. After controlling for age, sex, emotional instability, and quality of the relationship, significant effects in favor of the intervention group were found for self-efficacy care management ($F_{1,60}$=10.37; *P*=.002, *d*=0.85), and self-efficacy service use ($F_{1,60}$=11.47; *P*=.001; *d*=0.93), but not for depression ($F_{1,60}$=1.13; *P*=.293). Significant effects in favor of the intervention group were also demonstrated for mastery ($F_{1,60}$=12.66; *P*=.001; *d*=0.94), and quality of life ($F_{1,60}$=4.83; *P*=0.032; *d*=0.58), but not for perceived stress ($F_{1,60}$=3.40; *P*=0.071), and anxiety ($F_{1,60}$=0.80; *P*=.374).

Figure 3. Consolidated Standards of Reporting Trials (CONSORT) study flowchart.

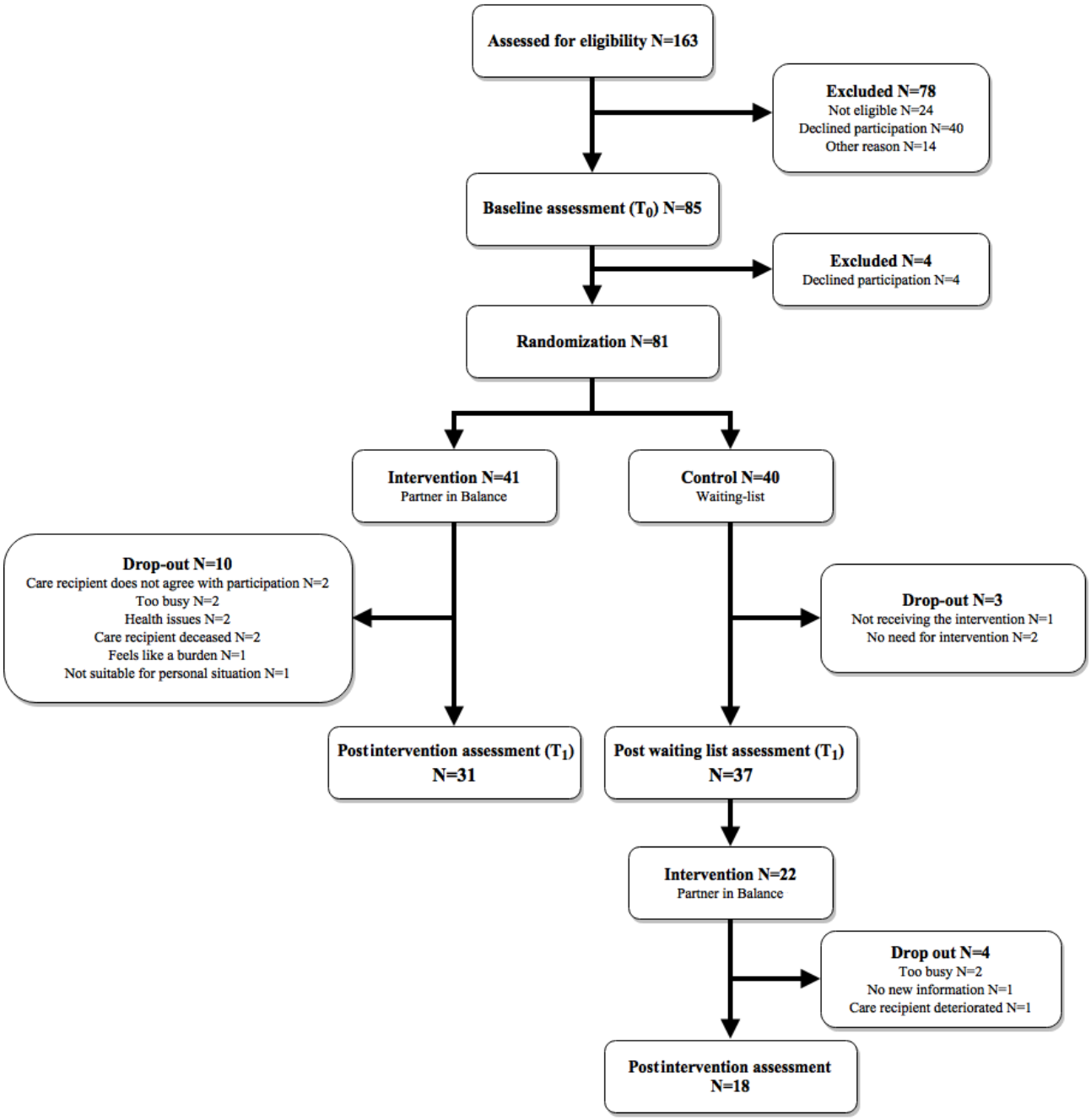


Table 1. Descriptive data for caregivers of both groups at baseline.

Demographics and outcome	Intervention (N=41)	Waiting list (N=40)	Test value comparing groups at baseline	*P* value
Socio-demographics				
Age, mean (SD)	67.8 (10.2)	70.2 (10.1)	1.0[a]	.302
Spouse, n (%)	37 (90.2)	37 (92.5)	0.6[b]	.432
Same household as PwD[c], n (%)	39 (95.1)	37 (92.5)	1.3[b]	.513
Care intensity in years, mean (SD)	1.8 (1.8)	1.9 (1.8)	674.5[d]	.929
Female, n (%)	29 (70.7)	24 (60.0)	0.8[b]	.385
Education, n (%)				
High school	8 (19.5)	4 (10.0)	2.3[b]	.321
College	18 (43.9)	16 (40.0)	—	—
Graduate school	15 (36.6)	20 (50.0)	—	—
Primary Outcomes				
Self-efficacy (CSES[e]), mean (SD)				
Care management	34.7 (7.8)	33.0 (9.4)	–0.9[a]	.395
Service use	25.8 (6.3)	23.7 (6.2)	–1.5[a]	.141
Depression (CES-D[f]), mean (SD)	13.1 (8.7)	13.1 (9.0)	732.0[d]	.927
Secondary Outcomes				
Stress (PSS[g]), mean (SD)	11.8 (6.0)	13.5 (6.2)	1.2[a]	.223
Anxiety (HADS-A[h]), mean (SD)	6.0 (3.7)	6.7 (4.7)	717.5[d]	.666
Mastery (PMS[i]), mean (SD)	23.7 (4.1)	22.9 (4.4)	–0.8[a]	.430
Quality of life (ICECAP-O[j]), mean (SD)	0.8 (0.1)	0.8 (0.1)	755.0[d]	.956

[a]Refers to *t* test (t_{79}).

[b]Refers to Chi-square test (χ^2_1).

[c]PwD: person with dementia.

[d]Refers to Mann-Whitney *U* Wilcoxon test.

[e]CSES: Caregiver Self-Efficacy Scale.

[f]CES-D: Center for Epidemiological Studies Depression Scale.

[g]PSS: Perceived Stress Scale.

[h]HADS-A: Hospital Anxiety and Depression Scale-Anxiety.

[i]PMS: Pearlin Mastery Scale.

[j]ICECAP-O: Investigating Choice Experiments for the Preferences of Older People.

Table 2. Analysis of covariance comparing intervention (N=31) and control (N=37) group at posttest.

	Control, mean (SD)		Intervention, mean SD		Mean difference[b] (95% CI)	*F* test (*df*)	Cohen *d*
Outcome	Crude[a]	Adjusted[b]	Crude	Adjusted			
Primary outcomes							
Self-efficacy (CSES[c])							
Care management	31.38 (8.71)	31.65 (1.05)	37.03 (6.33)	36.73 (1.12)	–5.07 (–8.23 to –1.92)	10.37[d] (1,60)	0.85
Service use	21.88 (6.33)	22.48 (0.83)	27.43 (5.11)	26.76 (0.89)	–4.27 (–6.80 to –1.75)	11.47[d] (1,60)	0.93
Depression (CES-D[e])	13.27 (9.21)	12.87 (1.08)	10.73 (8.20)	11.17 (1.14)	1.70 (–1.51 to 4.91)	1.13 (1,60)	0.28
Secondary outcomes							
Mastery (PMS[f])	21.15 (4.49)	21.32 (0.63)	24.87 (4.09)	24.68 (0.67)	–3.36 (–5.26 to –1.47)	12.66 (1,60)[d]	0.94
Stress (PSS[g])	13.76 (6.84)	12.92 (0.69)	10.03 (6.35)	10.99 (0.74)	1.94 (–0.17 to 4.04)	3.40 (1,60)	0.50
Anxiety (HADS-A[h])	5.94 (4.59)	5.91 (0.61)	6.70 (4.65)	6.73 (0.63)	–0.81 (–2.63 to 1.00)	0.80 (1,60)	0.24
Quality of life (ICECAP-O[i])	0.76 (0.15)	0.76 (0.02)	0.82 (0.10)	0.83 (0.02)	–0.06 (–0.12 to –0.01)	4.83 (1,60)[j]	0.58

[a]Group means.

[b]Adjusted for outcome measure at baseline, age, sex, education, quality of the relationship at baseline, neurotic personality traits, and coach background.

[c]CSES: Caregiver Self-Efficacy Scale.

[d]*P*<.01

[e]CES-D: Center for Epidemiological Studies Depression Scale.

[f]PMS: Pearlin Mastery Scale.

[g]PSS: Perceived Stress Scale.

[h]HADS-A: Hospital Anxiety and Depression Scale-Anxiety.

[i]ICECAP-O: Investigating Choice Experiments for the Preferences of Older People.

[j]*P*<.05

Discussion

Principal Findings

This randomized controlled study evaluated the first blended-care intervention for caregivers of people with early-stage dementia developed together with potential users, following the MRC Framework, and demonstrated a significant improvement in care management self-efficacy, service use self-efficacy, mastery, and quality of life after receiving the PiB intervention; compared to a waiting-list control group receiving care as usual. Effect sizes were medium (>0.5) for quality of life to high (>0.8) for self-efficacy and mastery. No differences between groups were demonstrated for caregiver depression, anxiety, and perceived stress.

Results on caregiver self-efficacy, mastery, and quality of life are in line with previous results in an uncontrolled study [25] and results of previous eHealth interventions for dementia caregivers [22]. Furthermore, the results of the present study fit the Stress and Coping paradigm by Lazarus and Folkman [8] and the Social Learning theory by Bandura [9], suggesting that taking charge of the changes in one's life increases self-efficacy and general wellbeing. Learning to positively manage life with dementia instead of managing the dementia itself in a self-management program may have facilitated caregivers' adaptation to their new caregiving role. The program's focus on enhancing positive, intact experiences that are tailored to the individual caregiver's situation could explain the positive effects on caregiver self-efficacy [11]. In addition, the relationship between the participant and the coach may have influenced the outcomes. The process evaluation of the present study showed that both participants and coaches mentioned that their relationship with each other had deepened [38], which was also demonstrated in a previous blended-care intervention for depression [18]. The opportunity to reflect on one's feelings anonymously in one's personal safe environment is easier than face-to-face, but the face-to-face contact increased caregiver openness, and therefore coach empathy with their situation [19]. However, we expected that higher levels of wellbeing or quality of life could be the result of a decrease in stress [8,9], which could not be derived from the results of the present study. It is conceivable that interventions aimed at the early stages may not be capable to decrease burden and stress, as these are relatively low during the early stages [7], leaving little room for improvement. Previous caregiver interventions demonstrating positive effects on burden and stress were not specifically aimed at early-stages of dementia [20,39-41]. The process evaluation also revealed that the intervention period and dose varied between participants. Moreover, the discussion forum was not used because caregivers mentioned that sharing their story felt like a betrayal to the care recipient and reading about other people's "misery" was considered undesirable [38]. These process characteristics may have influenced the intervention effectiveness [42]. Future follow up of PiB effects could clarify if an increase in self-efficacy results in a decrease

or prevention of increased stress and depression on the long term.

Strengths and Limitations

High face validity was demonstrated as the program was evaluated in multiple institutions with multiple coaches of different backgrounds. Development together with the potential users and a pilot evaluation following the MRC Framework may have increased its effectiveness.

The waiting-list period may have affected the differences in outcomes between both groups. The effects of waiting are highly variable and depend on the characteristics of the sample and of the trial [43]. However, this design allowed all potentially interested participants to participate in the intervention program, which may have increased their motivation to participate given that usual care for mild dementia caregivers often either does not include counseling or includes only infrequent counseling [44]. Furthermore, the waiting-list group was not deprived of usual care. An alternative would be a pseudo-intervention in which only psycho-education or only attention of the coach is provided, but the aim of this study was not to evaluate merely the online aspect of the intervention, but the effect of the blended-care intervention of which psycho-education and face-to-face contacts are integral parts.

Intention-to-treat analyses was not fully possible, as intervention noncompleters refused to participate in further assessments. However, we did include participants that were not completely compliant (completed only 2, 3 or no modules at all) in the analyses [38]. Drop-out was higher in the intervention group compared to the control group, which could have resulted in inflated effect sizes. However, selective drop-out was not demonstrated as completers did not differ from noncompleters at baseline. Often mentioned reasons for drop-out were no need for help or refusal by the care recipient, which was demonstrated previously as reasons of nonuse of formal services [45,46]. Furthermore, a higher rate of drop-out in the intervention group has previously been reported. Previous randomized controlled trials even controlled for any possible loss of power beforehand by increasing the sample of the intervention group. Nevertheless, the current effect sizes should be interpreted with caution. Although the power of our group was not jeopardized based on our power calculation, future studies could consider controlling for a higher rate of drop-out in the intervention group to prevent loss of power.

Our sample was not limited to memory clinics only, but the included participants may represent a subgroup of all dementia caregivers in the early stages. Caregivers in the early stages often decline formal care and it is conceivable that many were not familiar with the care parties involved in recruitment and were therefore overlooked in this study [45,46]. This could have resulted in a highly motivated sample more open to support [47]. Furthermore, only computer-literate caregivers could be included, which represents only around 59% of dementia caregivers [48]. However, seniors' use of internet is expected to rise in the near future [49], increasing the accessibility of PiB.

Future Research and Clinical Implications

Future research could consider combining all resources used during the intervention period with the intervention costs and outcomes in a cost-consequence analysis to aid decision makers. Furthermore, future research should evaluate sustainability of improvements at long-term follow-up. The higher rate of drop-out in the intervention group showed that this group feels overwhelmed but is perhaps most in need of the intervention. Some eHealth interventions show dropout rates of up to 80% [50-52] and therefore suggest blending face-to-face contacts with online modules, like the PiB program, to prevent these high drop-out rates. We found a relatively high response and participation rate [18], indicating that there is a need for at least having the option to choose for this type of caregiver support.

Conclusion

In conclusion, this study showed that a blended care self-management program for dementia caregivers in the early stages is effective in increasing caregiver self-efficacy, mastery, and quality of life on the short-term. The program could provide accessible care to future generations of caregivers of people with early-stage dementia and strengthen the primary caregivers.

Acknowledgments

This study is made possible by the financial support obtained from Alzheimer Nederland (Grant #WE03-2010-08) and the Alzheimer Research Fund Limburg. The authors would like to thank the family caregivers who participated in this study and the professional caregivers of MUMC+, Hulp bij Dementie, Virenze Riagg, MetGGZ, Steunpunt Mantelzorg, Catharina Hospital, and Elkerliek Hospital for their help in recruitment. Further, we thank Claudia Smeets, Inge Klinkenberg, Hanneke Withagen, and Caroline Sladek for their contribution in recruiting participants and performing the pre- and postintervention assessments.

Authors' Contributions

MEdV developed the project proposal and obtained funding. LMMB, MEdV, GIJMK, and FRJV designed the trial and the materials. LMMB monitored data collection for the whole trial, wrote the statistical analysis plan, cleaned and analyzed the data, and drafted and revised the paper. She is guarantor. MEdV and FRJV supervised the data collection and analysis. MEdV, GIJMK, and FRJV reviewed and revised the draft paper. All authors approved the final version of the paper to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 875KB - jmir_v20i7e10017_app1.pdf]

References

1. Alzheimer's Disease International. London: ADI; 2013. Policy Brief for Heads of Government: The Global Impact of Dementia 2013-2050 URL: https://www.alz.co.uk/research/GlobalImpactDementia2013.pdf [accessed 2018-04-26] [WebCite Cache ID 6yxpjjH6h]
2. de Vugt ME, Stevens F, Aalten P, Lousberg R, Jaspers N, Winkens I, et al. Do caregiver management strategies influence patient behaviour in dementia? Int J Geriatr Psychiatry 2004 Jan;19(1):85-92. [doi: 10.1002/gps.1044] [Medline: 14716704]
3. Brodaty H, Donkin M. Family caregivers of people with dementia. Dialogues Clin Neurosci 2009;11(2):217-228 [FREE Full text] [Medline: 19585957]
4. Olazarán J, Reisberg B, Clare L, Cruz I, Peña-Casanova J, Del ST, et al. Nonpharmacological therapies in Alzheimer's disease: a systematic review of efficacy. Dement Geriatr Cogn Disord 2010;30(2):161-178 [FREE Full text] [doi: 10.1159/000316119] [Medline: 20838046]
5. Gaugler JE, Kane RL, Kane RA, Newcomer R. Early community-based service utilization and its effects on institutionalization in dementia caregiving. Gerontologist 2005 Apr;45(2):177-185. [Medline: 15799982]
6. Kjallman Alm A, Hellzen O, Norbergh K. Experiences of long term ongoing structured support in early stage of dementia - a case study. Int J Older People Nurs 2014 Dec;9(4):289-297. [doi: 10.1111/opn.12034] [Medline: 23758956]
7. de Vugt ME, Verhey FRJ. The impact of early dementia diagnosis and intervention on informal caregivers. Prog Neurobiol 2013 Nov;110:54-62. [doi: 10.1016/j.pneurobio.2013.04.005] [Medline: 23689068]
8. Lazarus R, Folkman S. Stress, Appraisal, and Coping. New York: Springer Publishing Co Inc; 1984.
9. Bandura A. Self-Efficacy: The Exercise of Control. New York: Worth Publishers, Incorporated; 1997.
10. Acton GJ, Kang J. Interventions to reduce the burden of caregiving for an adult with dementia: a meta-analysis. Res Nurs Health 2001 Oct;24(5):349-360. [Medline: 11746065]
11. Boots LMM, Wolfs CAG, Verhey FRJ, Kempen GIJM, de Vugt ME. Qualitative study on needs and wishes of early-stage dementia caregivers: the paradox between needing and accepting help. Int Psychogeriatr 2015 Jun;27(6):927-936. [doi: 10.1017/S1041610214002804] [Medline: 25566686]
12. Macdonald A, Cooper B. Long-term care and dementia services: an impending crisis. Age Ageing 2007 Jan;36(1):16-22 [FREE Full text] [doi: 10.1093/ageing/afl126] [Medline: 17175565]
13. Lewis BA, Williams DM, Neighbors CJ, Jakicic JM, Marcus BH. Cost Analysis of Internet vs. Print Interventions for Physical Activity Promotion. Psychol Sport Exerc 2010 May 1;11(3):246-249 [FREE Full text] [doi: 10.1016/j.psychsport.2009.10.002] [Medline: 20401164]
14. Kirsch SE, Lewis FM. Using the World Wide Web in health-related intervention research. A review of controlled trials. Comput Inform Nurs 2004;22(1):8-18. [Medline: 15069844]
15. Harrell KM, Wilkins SS, Connor MK, Chodosh J. Telemedicine and the evaluation of cognitive impairment: the additive value of neuropsychological assessment. J Am Med Dir Assoc 2014 Aug;15(8):600-606. [doi: 10.1016/j.jamda.2014.04.015] [Medline: 24913209]
16. Serafini JD, Damianakis T, Marziali E. Clinical practice standards and ethical issues applied to a virtual group intervention for spousal caregivers of people with Alzheimer's. Soc Work Health Care 2007;44(3):225-243. [doi: 10.1300/J010v44n03_07] [Medline: 17548277]
17. Robinson L, Brittain K, Lindsay S, Jackson D, Olivier P. Keeping In Touch Everyday (KITE) project: developing assistive technologies with people with dementia and their carers to promote independence. Int Psychogeriatr 2009 Jun;21(3):494-502. [doi: 10.1017/S1041610209008448] [Medline: 19193255]
18. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014;14(1):355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
19. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]
20. Hayden LJ, Glynn SM, Hahn TJ, Randall F, Randolph E. The use of Internet technology for psychoeducation and support with dementia caregivers. Psychol Serv 2012 May;9(2):215-218. [doi: 10.1037/a0027056] [Medline: 22662739]
21. 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;13:17 [FREE Full text] [doi: 10.1186/1471-244X-13-17] [Medline: 23305463]
22. Boots LMM, de Vugt VME, van Knippenberg RJM, Kempen GIJM, Verhey FRJ. 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]

23. Boots LMM, de Vugt ME, Kempen GIJM, Verhey FRJ. Effectiveness of the blended care self-management program "Partner in Balance" for early-stage dementia caregivers: study protocol for a randomized controlled trial. Trials 2016 May 04;17(1):231 [FREE Full text] [doi: 10.1186/s13063-016-1351-z] [Medline: 27142676]
24. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr 1997;9 Suppl 1:173-176. [Medline: 9447441]
25. Boots LM, de Vugt ME, Withagen HE, Kempen GI, Verhey FR. Development and Initial Evaluation of the Web-Based Self-Management Program "Partner in Balance" for Family Caregivers of People With Early Stage Dementia: An Exploratory Mixed-Methods Study. JMIR Res Protoc 2016 Mar 01;5(1):e33 [FREE Full text] [doi: 10.2196/resprot.5142] [Medline: 26932438]
26. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [Medline: 12867348]
27. Fortinsky RH, Kercher K, Burant CJ. Measurement and correlates of family caregiver self-efficacy for managing dementia. Aging Ment Health 2002 May;6(2):153-160. [doi: 10.1080/13607860220126763] [Medline: 12028884]
28. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
29. Pearlin LI, Schooler C. The structure of coping. J Health Soc Behav 1978 Mar;19(1):2-21. [Medline: 649936]
30. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
31. Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? Soc Sci Med 2006 Apr;62(8):1891-1901. [doi: 10.1016/j.socscimed.2005.08.023] [Medline: 16168542]
32. Coast J, Flynn TN, Natarajan L, Sproston K, Lewis J, Louviere JJ, et al. Valuing the ICECAP capability index for older people. Soc Sci Med 2008 Sep;67(5):874-882. [doi: 10.1016/j.socscimed.2008.05.015] [Medline: 18572295]
33. Reisberg B, Ferris SH, de Leon MJ, Crook T. Global Deterioration Scale (GDS). Psychopharmacol Bull 1988;24(4):661-663. [Medline: 3249768]
34. Lawrence RH, Tennstedt SL, Assmann SF. Quality of the caregiver--care recipient relationship: does it offset negative consequences of caregiving for family caregivers? Psychol Aging 1998 Mar;13(1):150-158. [Medline: 9533197]
35. Costa P, McCrae R. NEO PI-R: Professional manual. Odessa, FL: Psychological Assessment Resources; 1992.
36. 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]
37. Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1977.
38. Boots LM, de Vugt ME, Smeets CM, Kempen GI, Verhey FR. Implementation of the Blended Care Self-Management Program for Caregivers of People With Early-Stage Dementia (Partner in Balance): Process Evaluation of a Randomized Controlled Trial. J Med Internet Res 2017 Dec 19;19(12):e423 [FREE Full text] [doi: 10.2196/jmir.7666] [Medline: 29258980]
39. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J, et al. Web-Based STAR E-Learning Course Increases Empathy and Understanding in Dementia Caregivers: Results from a Randomized Controlled Trial in the Netherlands and the United Kingdom. J Med Internet Res 2015;17(10):e241 [FREE Full text] [doi: 10.2196/jmir.4025] [Medline: 26519106]
40. Blom MM, Zarit SH, Groot ZRBM, 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]
41. 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]
42. Moncher FJ, Prinz RJ. Treatment fidelity in outcome studies. Clinical Psychology Review 1991 Jan;11(3):247-266. [doi: 10.1016/0272-7358(91)90103-2]
43. Hesser H, Weise C, Rief W, Andersson G. The effect of waiting: A meta-analysis of wait-list control groups in trials for tinnitus distress. J Psychosom Res 2011 Apr;70(4):378-384. [doi: 10.1016/j.jpsychores.2010.12.006] [Medline: 21414459]
44. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [Medline: 10987780]
45. Brodaty H, Thomson C, Thompson C, Fine M. Why caregivers of people with dementia and memory loss don't use services. Int J Geriatr Psychiatry 2005 Jun;20(6):537-546. [doi: 10.1002/gps.1322] [Medline: 15920707]
46. Wolfs CAG, de Vugt ME, Verkaaik M, Haufe M, Verkade P, Verhey FRJ, et al. Rational decision-making about treatment and care in dementia: a contradiction in terms? Patient Educ Couns 2012 Apr;87(1):43-48 [FREE Full text] [doi: 10.1016/j.pec.2011.07.023] [Medline: 21889285]
47. Banningh LWAJ, Vernooij-Dassen MJFJ, Vullings M, Prins JB, Rikkert MGMO, Kessels RPC. Learning to live with a loved one with mild cognitive impairment: effectiveness of a waiting list controlled trial of a group intervention on significant

others' sense of competence and well-being. Am J Alzheimers Dis Other Demen 2013 May;28(3):228-238. [doi: 10.1177/1533317513481093] [Medline: 23528880]
48. Kim H. Understanding Internet Use Among Dementia Caregivers: Results of Secondary Data Analysis Using the US Caregiver Survey Data. Interact J Med Res 2015;4(1):e1 [FREE Full text] [doi: 10.2196/ijmr.3127] [Medline: 25707033]
49. Eastman JK, Iyer R. The elderly's uses and attitudes towards the Internet. Journal of Consumer Marketing 2004 May;21(3):208-220. [doi: 10.1108/07363760410534759]
50. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
51. Geraghty AWA, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international Internet health interventions: a step towards global eHealth. Health Promot Int 2013 Sep;28(3):442-452 [FREE Full text] [doi: 10.1093/heapro/das029] [Medline: 22786673]
52. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

Abbreviations

ANCOVA: analysis of covariance
CES-D: Center for Epidemiological Studies Depression Scale
CSES: Caregiver Self-Efficacy Scale
eHealth: electronic health
HADS-A: Hospital Anxiety and Depression Scale-Anxiety
ICECAP-O: Investigating Choice Experiments for the Preferences of Older People
MRC: Medical Research Council
MUMC+: Maastricht University Medical Center
PiB: Partner in Balance
PMS: Pearlin Mastery Scale
PSS: Perceived Stress Scale
PwD: person with dementia

Edited by G Eysenbach; submitted 02.02.18; peer-reviewed by G Baumblatt, D Hansen; comments to author 15.03.18; revised version received 27.04.18; accepted 08.05.18; published 13.07.18.

Please cite as:
Boots LMM, de Vugt ME, Kempen GIJM, Verhey FRJ
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;20(7):e10017
URL: http://www.jmir.org/2018/7/e10017/
doi:10.2196/10017
PMID:30006327

Original Paper

Developing a Digital Marketplace for Family Planning: Pilot Randomized Encouragement Trial

Eric P Green[1], BA, MA, PhD; Arun Augustine[1], BS, MSc; Violet Naanyu[2], PhD; Anna-Karin Hess[1]; Lulla Kiwinda[1]

[1]Duke Global Health Institute, Duke University, Durham, NC, United States

[2]Department of Behavioral Sciences, School of Medicine, College of Health Sciences, Moi University, Eldoret, Kenya

Corresponding Author:
Eric P Green, BA, MA, PhD
Duke Global Health Institute
Duke University
Box 90519
Durham, NC, 27708
United States
Phone: 1 9196817289
Fax: 1 9196817748
Email: eric.green@duke.edu

Abstract

Background: Family planning is an effective tool for preventing death among women who do not want to become pregnant and has been shown to improve newborn health outcomes, advance women's empowerment, and bring socioeconomic benefits through reductions in fertility and population growth. Yet among the populations that would benefit the most from family planning, uptake remains too low. The emergence of digital health tools has created new opportunities to strengthen health systems and promote behavior change. In this study, women with an unmet need for family planning in Western Kenya were randomized to receive an encouragement to try an automated investigational digital health intervention that promoted the uptake of family planning.

Objective: The objectives of the pilot study were to explore the feasibility of a full-scale trial—in particular, the recruitment, encouragement, and follow-up data collection procedures—and to examine the preliminary effect of the intervention on contraception uptake.

Methods: This pilot study tested the procedures for a randomized encouragement trial. We recruited 112 women with an unmet need for family planning from local markets in Western Kenya, conducted an eligibility screening, and randomized half of the women to receive an encouragement to try the investigational intervention. Four months after encouraging the treatment group, we conducted a follow-up survey with enrolled participants via short message service (SMS) text message.

Results: The encouragement sent via SMS text messages to the treatment group led to differential rates of intervention uptake between the treatment and control groups; however, uptake by the treatment group was lower than anticipated (19/56, 33.9% vs 1/56, 1.8%, in the control group). Study attrition was also substantial. We obtained follow-up data from 44.6% (50/112) of enrolled participants. Among those in the treatment group who tried the intervention, the instrumental variables estimate of the local average treatment effect was an increase in the probability of contraceptive uptake of 41.0 percentage points (95% uncertainty interval −0.03 to 0.85).

Conclusions: This randomized encouragement design and study protocol is feasible but requires modifications to the recruitment, encouragement, and follow-up data collection procedures.

Trial Registration: ClinicalTrials.gov NCT03224390; https://clinicaltrials.gov/ct2/show/NCT03224390 (Archived by WebCite at http://www.webcitation.org/70yitdJu8)

(J Med Internet Res 2018;20(7):e10756) doi:10.2196/10756

KEYWORDS

family planning; unmet need; contraception; digital health; Kenya

Introduction

Family planning is one of the most effective public health interventions, and more women than ever before are experiencing the benefits. Voluntary family planning has been shown to prevent maternal death among women who do not want to become pregnant [1], improve newborn health outcomes [2], and bring socioeconomic benefits through reductions in fertility and population growth [3]. Contraceptive use may also advance women's empowerment, but the evidence is weak [4,5]. Globally, in 2017, an estimated 715 million married or in-union women of reproductive age were using a modern method of contraception (58%), an increase of 22% since 2000 [6].

Despite this positive trend, another 203 million women—primarily in Asia and Africa—want to prevent or delay pregnancy but are not using a modern method of contraception [6]. This situation is referred to as an *unmet need* for modern contraception, and it signals the presence of barriers to uptake that may include limited access to methods, concerns about side effects, and other issues such as cultural norms against use. The proportion of women with an unmet need for contraception is highest in Africa, where more than 46 million married and in-union women (22%) would like to prevent or delay childbirth but are not using a modern method of contraception.

In Kenya, for instance, 17% of currently married and in-union women of reproductive age [7] and 26% of sexually active unmarried women [8] have an unmet need for family planning. This translates into approximately 1.3 million women in the country who are not using contraception but say they would like to avoid pregnancy. Millions of others are either unaware of the potential benefits of contraception, misinformed about the full range of modern methods available, or unsatisfied with previous experiences using contraception [7].

In recognition of the needs of women and girls living in Kenya and beyond, a major international initiative called Family Planning 2020 launched at the London Summit on Family Planning with the goal of "expanding access to family planning information, services, and supplies to an additional 120 million women and girls in 69 of the world's poorest countries by 2020." This initiative has sparked important gains, but more work remains if this goal is to be realized. Since the launch of FP2020 in 2012, an additional 38.8 million women have begun using a modern method of contraception [9]. While this progress is above historic trends, it is substantially off the pace required to meet the goal of adding 120 million new users by 2020. This gap suggests the need for new approaches that can augment existing efforts to expand the coverage of family planning.

Traditionally, efforts to promote the uptake of family planning have focused on demand generation activities, supply-side activities, or a mixture of both. Demand generation interventions seek to change knowledge, attitudes, and practices regarding family planning. Common approaches include mass media advertising (also known as behavior change communication), one-on-one and small group discussions, and economic incentives such as conditional cash transfer programs. Supply-side interventions aim to increase access, improve quality, and lower costs for family planning services. A systematic review of 63 published evaluations of family planning interventions concluded that economic incentives and supply-side interventions had the most consistent effect on contraceptive use, but the overall quality of the evidence was low [10].

The emergence of digital health tools—such as short message service (SMS), interactive voice response, and mobile phone apps—have created new opportunities to strengthen health systems and promote behavior change [11,12], but the evidence base for digital health remains weak. As is the case for nondigital interventions [10], studies of digital health tools have found that it is easier to increase knowledge than to change behavior [13].

This pilot study represents another effort to promote behavior change through the use of an SMS text messaging intervention. Women with an unmet need for modern methods of contraception in Western Kenya were randomized to receive messages that encouraged them to try an investigational digital health intervention. The objectives of this pilot study were to explore the feasibility of a full-scale trial—in particular the recruitment, encouragement, and follow-up data collection procedures—and to examine the preliminary effect of the intervention on the uptake of contraception.

Methods

Recruitment

This was an external pilot study [14,15] conducted to inform the design and implementation of a separate full-scale trial. The study design was a randomized encouragement trial.

Setting and Participants

The target population for this study was Kenyan women who had an unmet need for family planning, that is, women who were not using family planning but wished to delay or prevent pregnancy. The accessible population was limited to women with an unmet need living in Bungoma County, Kenya.

Recruitment and Eligibility Screening

Over a period of 4 weeks in 2017, from July 12 to August 6, we conducted recruitment exercises at 6 open-air markets throughout the county. We identified 21 market venues in Bungoma County and selected 5 large markets and 1 small market that maximized geographical coverage. We visited each market on its "market day," the day of the week when foot traffic peaks. Market days for the selected markets were Sunday, Monday, Wednesday, Thursday, and Friday. Our team visited 2 markets on Fridays.

Our market stall advertised an opportunity to participate in the "Bungoma County Women's Health Study." A team of 4 female study team members, all Kenyan, staffed the study table and screened women for eligibility. To be eligible to enroll in the study, women had to (1) be between the ages of 18 and 35 years (inclusive), (2) have an unmet need for family planning, (3) live in Bungoma County, (4) demonstrate phone ownership, (5) opt-in to receiving calls and SMS text messages related to the study, (6) demonstrate basic ability to operate the study tablets,

and (7) provide consent to participate in the study. Women who were pregnant or fewer than 4 months postpartum were excluded.

To begin the screening with an interested woman, a member of the study team asked the woman her age and county of residence. To demonstrate phone ownership and continue to the second stage of screening, the woman had to show the enumerator that she received a test SMS text message from the study shortcode. In the second phase of screening, the enumerator asked the woman if she was pregnant or currently using any method of family planning to prevent or delay pregnancy.

If the woman was eligible to move to the third stage of screening, the enumerator demonstrated how to use the tablet computer to complete the survey via audio computer assisted self-interview. The screening survey text and audio were available in English and Swahili. The woman had to demonstrate proficiency in an example exercise to continue to the full screening. Enumerators were on hand to assist participants who needed help using the tablet.

Unmet Need

In the third and final phase of screening, the woman completed the baseline survey to enable us to classify her unmet need status and to collect relevant background information. The baseline survey instrument included several modules from the 2014 Kenya Demographic and Health Survey (Phase 7, short form), including household characteristics, respondent's background, reproduction, contraception, and marriage and sexual activity [7].

To define unmet need for this study, we followed guidelines published by the Demographic and Health Survey Program (DHS, revised 2012) [16] and other relevant scholarly reviews [17]. A woman was classified as having an unmet need if she reported no current use of contraception, was not identified by the survey as infecund, and said she did not want to be pregnant for at least 2 years. A woman could also be classified as having an unmet need if she was postpartum amenorrheic and reported that she did not want her last birth at all or wanted to become pregnant later than she did. We further classified women as having an unmet need for limiting (does not want to become pregnant at all) or spacing (wants to delay pregnancy for at least 2 years). We extended this classification of unmet need to women who were not married or in a union if they reported being sexually active in the past 6 months, thus, putting them at risk for pregnancy. See our Multimedia Appendices for survey questions (Appendix 1) and a detailed algorithm for determining unmet need (Appendix 2).

Enrollment

If a woman was eligible to participate in the study based on her responses to the screening, the tablet prompted the enumerator to review the informed consent form with her. If she consented to participate, the enumerator recorded her name and contact details in the study register. Every woman who completed the screening received an honorarium of KES 200 (approximately US $2) for her time and effort, regardless of whether she was eligible to participate in the study or consented to participate. Ineligible women were not informed about the specific reason that they were ineligible to prevent others from determining which answers would trigger eligibility. At the time of enrollment, we informed participants that we might invite them to learn more about family planning and women's health with one of our partners.

Intervention

The investigational intervention was a digital health marketplace for family planning called Nivi [18]. At the time of the study, any woman (or man) in Bungoma County could send a toll-free SMS text message to the Nivi service to ask a question about reproductive health or trigger a free callback to complete an automated family planning counseling session via interactive voice response. This session resulted in a set of recommended methods that fit the client's preferences and goals, along with referrals to local public and private providers offering one or more of these methods. After a period of time, clients were prompted to provide details about their experience with family planning providers and were eligible to receive a transportation voucher (approximately USD $2) as a nudge toward behavior change. The investigational intervention remained under active development during the pilot trial. Participants could text or call customer service representatives as needed.

Experimental Design and Randomization

Since the service was available to anyone living in Bungoma County, it was not possible to restrict access and estimate the impact of the service through a randomized controlled trial. In situations like this, a randomized encouragement design can be very effective [19]. In a randomized encouragement design, participants are randomized to receive an invitation or special encouragement to receive an intervention. Not everyone who is encouraged will try the intervention (and some who are not invited will try it on their own); however, as long as those randomly assigned to receive the encouragement—"the treatment group"—try the intervention at a higher rate than those not encouraged—the "control group"—it is possible to estimate the impact of the intervention. This design has been used to study various interventions where two-sided noncompliance is possible [20-23].

In this pilot trial, we randomly allocated the sample of 112 enrolled women to the treatment or control arm (1:1). At the end of the recruitment period, the first author used the blockTools package [24] in R [25] to block randomize by age and baseline indicators of having attended postsecondary schooling, previous use and discontinuation of contraception, and being married or living in a union. One month after the end of the recruitment period, on October 2, 2017, women randomized to the encouragement arm received an invitation via SMS text message to try the service and complete a free family planning screening (plus bonus phone credit of approximately US $2, not conditional on the use of service). Women randomized to the control arm received a different set of messages thanking them for participating in the study; the control messages did not mention the investigational service.

Outcome Data Collection

We conducted a follow-up survey between February 14 and March 13, 2018, approximately 4 months after we invited the treatment group to try the service. Participants could complete the survey for free via an SMS text message in their preferred language or choose to receive a free callback from a study enumerator to complete the survey over the phone. Any woman who attempted to send an SMS text message but experienced an error was flagged for enumerator follow-up. The study enumerator was blind to each participant's assignment until the end of the survey. We sent up to 4 SMS text message reminders from our study shortcode (mask "DGHI") to study participants who did not reply. Women who completed the survey received an honorarium of KES 200 (approximately US $2) to appreciate their time and effort.

The primary outcome under investigation was self-reported use of a modern method of contraception [26] since the baseline survey. This included women who adopted and subsequently discontinued a method during this period. The reference point for the start of the recall period was the national election conducted on August 8, 2017, several days after the end of the baseline survey. We obtained a binary indicator of attempted service use by querying the system logs for participant phone numbers. If a participant's phone number was present in the system logs, we coded her as having tried the service.

Statistical Analysis

Because encouragement designs lead to two-sided noncompliance, we planned to use instrumental variables regression to obtain an unbiased local average treatment effect (LATE) of the impact of service use on contraceptive uptake. We used the AER [27] package in R [25] to estimate LATE via two-stage least squares regression. In the first stage, we regressed the indicator of service use on the instrumental variable—a binary indicator of random assignment to the treatment group. In the second stage, we regressed the primary outcome of contraceptive uptake on the predicted values of service use from the first stage regression. Both regressions included baseline controls and the mode of follow-up survey. We used the ivpack [28] package to obtain corrected Huber-White SEs. The results of nonlinear specifications are presented in Multimedia Appendix 3.

For this approach to be valid, the instrumental variable (or instrument) must meet 3 assumptions: (1) The instrument is randomly assigned (independence assumption), (2) the instrument increases use of the investigational intervention, and (3) the instrument only affects the outcome through use of the intervention (exclusion restriction) [29]. We satisfy the independence assumption (1) through the randomized design, and we demonstrate assumption (2) to be true empirically. There is not a direct test of the exclusion restriction (3), but it seems reasonable to assume this is met because the encouragement to try the service did not itself encourage women to adopt contraception or otherwise counsel them on the importance of family planning.

One aim of the study was to test the recruitment procedures and examine the potential for attrition. We based the target sample size for the full trial on the assumption that a sample size of 50 would be needed in an individually randomized trial (25 per arm) to detect a difference in contraception uptake of 30 percentage points between the control group (10%) and the treatment group (40%), given an alpha of 5%, power of 80%, and a one-tailed test. We increased this sample size estimate by a factor of 2.8 to account for the fact that only a subset of the treatment group was expected to uptake the intervention (70%) and that there would be a differential rate of service uptake in the control group that was not encouraged (10%). The inflation factor was $1/(0.7 - 0.1)^2$, producing an adjusted target sample size of 139 [30].

Ethical Review

Institutional Review Boards at Duke University and Moi University reviewed and approved this study protocol. This pilot study is registered with ClinicalTrials.gov (NCT03224390).

Results

Participant Characteristics

As shown in Figure 1, we assessed 772 women for eligibility and enrolled 112 women. A total of 660 women were excluded because they did not meet the inclusion criteria; 33.0% (218/660) of excluded women had a met need for contraception.

Table 1 summarizes the characteristics of the enrolled sample. The average age of participants was 24.7 (SD 4.8) years. The majority of women in the study were married or in a union, and two-thirds reported previous pregnancies. The average woman gave birth to 1.6 (SD 1.6) children and desired to have a total of 3.6 (SD 1.3) children. Most women reported an unmet need for spacing, rather than limiting. As is typical of women in Bungoma County, according to the most recent DHS, the women in this study were familiar with family planning methods. Most women indicated that they had recently been exposed to family planning messages in the media, and the average woman said she had heard of 9.6 (SD 2.2) out of 12 methods assessed.

Intervention Uptake

The randomized encouragement design had only a modest effect on the probability of trying the intervention. Four months after the treatment group was encouraged via SMS text message to try the service, 33.9% (19/56) of women in the treatment group initiated a session, compared with only 1.8% (1/56) in the control group. The encouragement did produce a differential rate of uptake of 32.1 percentage points, but the difference was smaller than anticipated.

Table 2 shows the correlates of intervention use among the treatment group. Age was negatively associated with use, which was expected. No other baseline characteristics of participants were significantly associated with use.

Figure 1. Participant flow diagram.

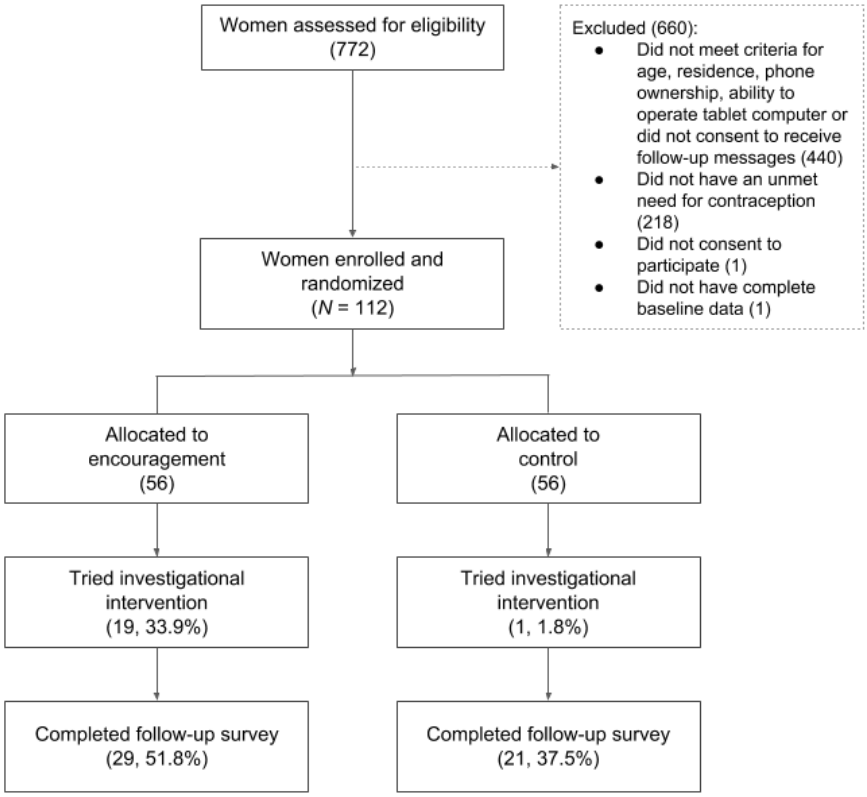


Table 1. Participant characteristics.

Characteristic	Control (N=56)	Treatment (N=56)	Kenya Demographic and Health Survey Program 2014 Reference	
			Value	Reference Group
Age in years, mean (SD)	24.9 (4.6)	24.6 (5.0)	N/A[a]	N/A
Married or in union, %	55.4	60.7	59.7	All women, national, 20-24 years
Christian, %	96.4	94.6	91.4	All women, national, 15-49 years
Luhya tribe, %	75.0	78.6	15.0	All women, national, 15-49 years
Attended postsecondary schooling, %	19.6	17.9	7.2	All women, Bungoma, 15-49 years
No schooling, %	3.6	0.0	0.9	All women, Bungoma, 15-49 years
Nulligravida, %	30.4	33.9	35.3	All women, national, 20-24 years
Number of children born, mean (SD)	1.7 (1.6)	1.5 (1.6)	1.1[b]	All women, national, 20-24 years
Number of desired children, mean (SD)	3.7 (1.4)	3.4 (1.1)	3.6[b]	All women, national, 15-49 years
Unmet need for spacing, %	78.6	82.1	90.5	Currently married women[c], national, 15-49 years
Past use of family planning, %	75.0	67.9	30.5	All women[d], national, 15-49 years
Number methods known, mean (SD)[e]	9.7 (1.9)	9.4 (2.5)	8.7[b]	All women, national, 15-49 years
Not exposed to family planning messages, %[f]	21.4	17.9	18.9	All women, Western, 15-49 years

[a]N/A: not applicable.

[b]Standard deviation not reported.

[c]Currently married women with an unmet need for family planning.

[d]Women who started an episode of contraceptive use within the 5 years preceding the survey and discontinued within 12 months.

[e]Asked about knowledge of 12 different methods.

[f]Did not hear or see a family planning message on a radio or television or read in a newspaper or magazine in the past few months.

Table 2. Correlates of intervention uptake.

Characteristic[a]	Dependent variable (tried intervention)	*P* value
Age, beta (SE)	−.04 (.02)	.06
Married or in a union, beta (SE)	−.16 (.20)	.44
Identifies as Christian, beta (SE)	.44 (.33)	.20
Identifies as a member of Luhya tribe, beta (SE)	−.01 (.18)	.97
Attended postsecondary schooling, beta (SE)	.20 (.17)	.25
Nulligravida, beta (SE)	.09 (.24)	.71
Number of children born, beta (SE)	.04 (.10)	.65
Desired number of children, beta (SE)	.06 (.09)	.47
Has unmet need for spacing, beta (SE)	−.03 (.24)	.91
Past use of family planning, beta (SE)	−.26 (.17)	.14
Number of methods known, beta (SE)	.01 (.03)	.75
Not exposed to family planning messages, beta (SE)	−.19 (.20)	.35
Constant, beta (SE)	.71 (.58)	.23
Mean of dependent variable	.34	N/A[c]
Observations	56	N/A
R^2	.28	N/A
Adjusted R^2	.08	N/A
Residual SE	$.46_{43}$	N/A
F statistic	$1.40_{12,43}$	N/A

[a]Sample limited to women randomly assigned to the treatment group.

[b]Coefficients estimated through linear probability model regression.

[c]N/A: not applicable.

Study Attrition

As shown in Figure 1, there was a substantial amount of attrition. We obtained follow-up data from 44.6% (50/112) of enrolled participants. Slightly more than half (56.0%, 28/50) of participants who completed the follow-up survey did so via an SMS text message (vs via a phone call with a study enumerator). Table 3 shows that attrition was higher among the control group, but this difference was not statistically significant. Attrition was significantly associated with a few baseline characteristics, including postsecondary education, nulligravida, and the mean number of children born; participants found at endline were more likely to have attended postsecondary schooling, have never been pregnant, and have fewer children. The impact analysis controls for these baseline characteristics and the mode of survey administration. Missing follow-up observations were imputed with baseline values (last observation carried forward), which in this study was no contraceptive use on study entry.

Effects of Intervention Use

Table 4 presents preliminary evidence of the impact of the investigational intervention on the adoption of contraception. We found that assignment to the treatment group (ie, assignment to receive an encouragement to try the intervention) led to an increase of 12.7 percentage points in the likelihood of contraception use. This is the reduced form estimate (ie, the effect of the invitation—encouragement—on the uptake of contraception). The causal effect of interest, however, is the ratio of the reduced form estimate to the first stage estimate: 41.0 percentage points. This effect is known as LATE, and it represents the average causal effect for women whose use of the intervention was determined only through the random encouragement to try the intervention. In other words, it is *the effect of using the intervention on contraceptive uptake.* The sign of this estimate appears to be positive, but the CI is wide.

Two additional specifications are presented in Multimedia Appendix 3: (1) ordinary least squares estimates produced without the use of last observation carried forward imputation for missing data and (2) probit regression estimates. In the models based on the subset of complete data (1), the estimates and CIs are slightly wider than those presented in Table 4. In the nonlinear specifications (2), the results are consistent with the linear results presented in Table 4.

Table 3. Baseline participant characteristics by follow-up status.

Characteristic	Not found (N=62)	Found (N=50)	*P* value[a]
Assigned to treatment, n (%)	27 (44)	29 (58)	.18
Age, mean (SD)	25.0 (5.0)	24.4 (4.6)	.54
Married or in union, n (%)	39 (63)	26 (52)	.33
Christian, n (%)	58 (94)	49 (98)	.50
Luhya tribe, n (%)	47 (76)	39 (78)	.96
Attended postsecondary schooling, n (%)	7 (11)	14 (28)	.045
No schooling, n (%)	2 (3)	0 (0)	.57
Nulligravida, n (%)	15 (24)	21 (42)	.07
Number of children born, mean (SD)	1.8 (1.7)	1.2 (1.3)	.04
Number of desired children, mean (SD)	3.7 (1.4)	3.4 (1.1)	.13
Unmet need for spacing, %	47 (76)	43 (86)	.27
Past use of family planning, n (%)	46 (74)	34 (68)	.61
Number methods known, mean (SD)[b]	9.4 (2.3)	9.8 (2.1)	.31
Not exposed to family planning messages[c], n (%)	14 (23)	8 (16)	.53

[a]Two-sample *t* tests of mean differences and two-proportions z tests of differences in proportions.

[b]Asked about knowledge of 12 different methods.

[c]Did not hear or see a family planning message on a radio or television or read in a newspaper or magazine in the past few months.

Table 4. Impact on contraception adoption (N=122).

Model details[a]	Tried intervention	Adopted contraception	
	First stage regression estimate[b] (95% CI)	Intent-to-treat estimate[c] (95% CI)	Instrumental variables estimate[d] (95% CI)
Assigned to treatment	.31 (0.19 to 0.44)	.13 (−0.01 to 0.26)	N/A[e]
Tried intervention	N/A	N/A	.41 (−0.03 to 0.85)
Mean in control group	.02	.16	N/A

[a]Models include the following controls: an indicator for mode of follow-up survey administration and several baseline characteristics, including age, number of children born, and indicators for having attended postsecondary schooling, past use of family planning, being married or in a union, and nulligravida.

[b]The first stage regression estimate is the coefficient on assignment to treatment from an ordinary least squares regression of intervention use on assignment.

[c]The intent-to-treat estimate is the coefficient on assignment to treatment from an ordinary least squares regression of contraception adoption on assignment.

[d]The instrumental variables estimate is the coefficient on intervention use in a two-stage least squares regression of contraception adoption on assignment and intervention use.

[e]N/A: not applicable.

Discussion

Principal Findings

This pilot study demonstrates that the proposed recruitment, encouragement, and data collection procedures are feasible, but some modifications are necessary prior to conducting a full trial. Additionally, analysis of the pilot data suggests that the investigational intervention may have a positive effect on contraceptive uptake among women with an unmet need in Kenya, but a full trial is required to replicate the direction of this effect and more precisely estimate the effect size.

During a recruitment period that lasted 4 weeks, we screened 772 women for eligibility, but only enrolled 14.5% (112/772) in the study. At this rate, it would have taken another week to reach our original target sample size. While this approach was feasible in terms of time and resources, it was inefficient in two ways. First, two-thirds of women who were ineligible to enroll did not meet the basic eligibility criteria such as age, residence, and phone ownership. Screening out these women was not time intensive, but we could have eliminated some work and inconvenience to interested women by more clearly stating the criteria on the market stall signage. Second, 1 out of every 3 ineligible women was ineligible because they did not have an unmet need for family planning. To some extent, this was

unavoidable because we did not directly recruit women with an unmet need, but rather embedded checks for eligibility in a short screening available to all women in the eligible age range. In a future trial, it may be advantageous to recruit from other subpopulations in addition to open-air markets to increase the probability that the pool of potential participants will have an unmet need. For instance, recruiting from postsecondary institutions would enable us to reach younger, unmarried women who may be sexually active but not using contraception. Postnatal clinics are another potential venue for recruitment as there is a high unmet need among new mothers in this region.

We used a randomized encouragement design to account for expected two-sided noncompliance with treatment assignment. Women assigned to the treatment group received an invitation via an SMS text message to try the intervention, and 33.9% of these women accepted the invitation, a conversion rate that appears to be consistent with SMS text message marketing conversion rates observed in industry [31]. By comparison, 1.8% of control participants tried the intervention. The encouragement led to a differential rate of intervention uptake of 32.1 percentage points, thereby making causal identification possible using assignment to treatment as an instrument.

The intervention uptake rate is important because incomplete uptake requires an inflation of sample size estimates that are based on fixed parameters for power, alpha, and the desired minimal detectable effect size for traditional randomized controlled trials. Another important consideration for the optimal sample size is attrition. In this study, 44.6% of enrolled participants completed the follow-up survey via an SMS text message or a phone call with a study enumerator. We did not collect detailed tracking information from participants during the recruitment process, so we could only invite participants to complete the survey via an SMS text message. In a future trial, it will be important to have the option to conduct in-person follow-up to reduce study attrition. Other studies that relied solely on SMS text message invitations as we did have encountered similar challenges [13].

A third key consideration for sample size calculations is the minimal detectable effect size. In this study, the instrumental variables estimate of the treatment effect was an increase in the likelihood of contraception uptake of 41.0 percentage points among the treatment group members who tried the intervention. This is an approximate standardized effect size of 1.1; however, this is only a point estimate, and 95% CI is wide. While the results suggest that the intervention effect may be positive, the point estimate is not measured precisely. The effect observed in this study is large relative to other SMS text message interventions for health behavior change [13,32,33], so it will be important to use a more conservative estimate to determine the optimal sample size for the full trial.

Limitations

The main limitation of this study was attrition. While attrition was not significantly associated with treatment assignment, found and unfound participants at endline differed on a few baseline characteristics. The preliminary impact analysis controls for these differences, but selection bias is a concern. Our reliance on self-reported data, while standard for a trial like this, also has the potential for bias.

As this study was conducted in only one, largely rural county in Kenya, the results may not generalize to urban or international markets. Additionally, the study was conducted at a unique and challenging time. A few days after the end of the recruitment period, Kenyans voted in a national election that was ultimately nullified by the Supreme Court. A second election took place on October 26, 2017, roughly 2 weeks after the treatment group was encouraged to try the intervention. Then in early November, a 5-month national nurse's strike came to an end, and nurses around the country—including the bulk of the country's family planning service providers—returned to work. In short, the pilot study was conducted during a period of uncertainty, likely distrust of SMS text message marketing amid heavy political advertising, and a significant decrease in the availability of family planning providers. Given these extenuating circumstances, we attempted to follow-up with participants approximately 4 months after the treatment group was encouraged to try the service rather than 1 month as originally planned.

Conclusions

This randomized encouragement design and study protocol is feasible but requires modifications to the recruitment, encouragement, and follow-up data collection procedures.

Acknowledgments

This research was supported by a pilot grant from the Duke Global Health Institute. For their assistance with data collection, the authors would like to thank Maximilla Chivini, Nancy Wairmu, Mercy Murungi, and Purity Nyangweso. The authors also appreciate feedback on study design from Dr Ben Bellows, Siddhartha Goyal, and the Nivi team.

Authors' Contributions

EPG and AA designed the study, conducted the analysis, and led the write-up. VN contributed to the study design and manuscript. AKH and LK assisted with data collection and contributed to the manuscript.

Conflicts of Interest

EPG is a co-founder of Nivi, Inc, holds an equity stake in the company, is a member of the company's Board of Directors, and serves as the company's Chief Scientist. EPG is a faculty member in the Duke Global Health Institute. Duke University also

holds an equity stake in the company. EPG's potential conflicts of interest are managed by Duke University's Research Integrity Office (MP#0600050-2017-001-A).

Multimedia Appendix 1

Survey instruments.

[PDF File (Adobe PDF File), 233KB - jmir_v20i7e10756_app1.pdf]

Multimedia Appendix 2

Algorithm for determining unmet need.

[PDF File (Adobe PDF File), 361KB - jmir_v20i7e10756_app2.pdf]

Multimedia Appendix 3

Supplemental tables.

[PDF File (Adobe PDF File), 110KB - jmir_v20i7e10756_app3.pdf]

Multimedia Appendix 4

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 511KB - jmir_v20i7e10756_app4.pdf]

References

1. Campbell OMR, Graham WJ, Lancet Maternal Survival Series steering group. Strategies for reducing maternal mortality: getting on with what works. Lancet 2006 Oct 07;368(9543):1284-1299. [doi: 10.1016/S0140-6736(06)69381-1] [Medline: 17027735]
2. Cleland J, Bernstein S, Ezeh A, Faundes A, Glasier A, Innis J. Family planning: the unfinished agenda. Lancet 2006 Nov 18;368(9549):1810-1827. [doi: 10.1016/S0140-6736(06)69480-4] [Medline: 17113431]
3. Bongaarts J, Cleland J, Townsend J, Bertrand J, Gupta M. Family Planning Programs For the 21st Century. New York: Population Council; 2012. URL: http://www.popcouncil.org/research/family-planning-programs-for-the-21st-century-rationale-and-design [accessed 2018-07-19] [WebCite Cache ID 7125iD5nj]
4. Prata N, Fraser A, Huchko MJ, Gipson JD, Withers M, Lewis S, et al. WOMEN'S EMPOWERMENT AND FAMILY PLANNING: A REVIEW OF THE LITERATURE. J Biosoc Sci 2017 Nov;49(6):713-743 [FREE Full text] [doi: 10.1017/S0021932016000663] [Medline: 28069078]
5. Silverman R, Birdsall N, Glassman A. Center for Global Development. 2016. Can access to contraception deliver for women?s economic empowerment? What we know ? and what we must learn URL: https://www.cgdev.org/sites/default/files/Silverman-Birdsall-Glassman-Access-Contraception-Women-2016.pdf [accessed 2018-04-05] [WebCite Cache ID 6yRQmHsop]
6. United N, Department OEA. United Nations. 2017. World Family Planning 2017 - Highlights URL: http://www.un.org/en/development/desa/population/publications/pdf/family/WFP2017_Highlights.pdf [WebCite Cache ID 6yRR3psKd]
7. Kenya NBOS, Kenya MOH, National AIDSCC, Kenya MRI, National CFP. Kenya Demographic and Health Survey. Nairobi, Kenya: KNBS; 2015. URL: https://dhsprogram.com/pubs/pdf/fr308/fr308.pdf [accessed 2018-07-19] [WebCite Cache ID 7125pj1ev]
8. Kenya NBOS, Kenya MOH, National AIDSCC, Kenya MRI, National CFP. Kenya Demographic and Health Survey 2014 Key Indicators. Kenya: KNBS; 2014. URL: https://tinyurl.com/y7c7l73q
9. FP2020. 2017. FP2020: The Way Ahead URL: http://progress.familyplanning2020.org/en [accessed 2018-04-05] [WebCite Cache ID 6yRSKH03Z]
10. Mwaikambo L, Speizer IS, Schurmann A, Morgan G, Fikree F. What works in family planning interventions: a systematic review. Stud Fam Plann 2011 Jun;42(2):67-82 [FREE Full text] [Medline: 21834409]
11. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
12. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17 Suppl 1:82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]

13. Johnson D, Juras R, Riley P, Chatterji M, Sloane P, Choi SK, et al. A randomized controlled trial of the impact of a family planning mHealth service on knowledge and use of contraception. Contraception 2017 Jan;95(1):90-97. [doi: 10.1016/j.contraception.2016.07.009] [Medline: 27421767]
14. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. Stat Med 1990;9(1-2):65-71; discussion 71. [Medline: 2345839]
15. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. J Eval Clin Pract 2004 May;10(2):307-312. [doi: 10.1111/j.2002.384.doc.x] [Medline: 15189396]
16. Bradley S, Croft T, Fishel J, Westoff C. Revising unmet need for family planning. Calverton, Maryland: ICF International and MEASURE DHS; 2012. URL: https://dhsprogram.com/pubs/pdf/AS25/AS25%5B12June2012%5D.pdf [accessed 2018-07-19] [WebCite Cache ID 7125y7U1d]
17. Bradley SEK, Casterline JB. Understanding unmet need: history, theory, and measurement. Stud Fam Plann 2014 Jun;45(2):123-150 [FREE Full text] [doi: 10.1111/j.1728-4465.2014.00381.x] [Medline: 24931072]
18. Nivi. 2016. URL: https://www.nivi.io/ [accessed 2018-04-05] [WebCite Cache ID 6yRSP9BR3]
19. West SG, Duan N, Pequegnat W, Gaist P, Des JDC, Holtgrave D, et al. Alternatives to the randomized controlled trial. Am J Public Health 2008 Aug;98(8):1359-1366. [doi: 10.2105/AJPH.2007.124446] [Medline: 18556609]
20. Duflo E, Saez E. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. The Quarterly Journal of Economics 2003;118(3):815-842.
21. Devoto F, Duflo E, Dupas P, Parienté W, Pons V. Happiness on Tap: Piped Water Adoption in Urban Morocco. American Economic Journal: Economic Policy 2012 Nov;4(4):68-99. [doi: 10.1257/pol.4.4.68]
22. Holland PW. Causal Inference, Path Analysis, and Recursive Structural Equations Models. Sociological Methodology 1988;18:449. [doi: 10.2307/271055]
23. Hirano K, Imbens GW, Rubin DB, Zhou XH. Assessing the effect of an influenza vaccine in an encouragement design. Biostatistics 2000 Mar;1(1):69-88. [doi: 10.1093/biostatistics/1.1.69] [Medline: 12933526]
24. Moore, Ryan T. BlockTools: Blocking, assignment, and diagnosing interference in randomized experiments. 2016. URL: https://cran.r-project.org/web/packages/blockTools/blockTools.pdf [accessed 2018-07-19] [WebCite Cache ID 7126QYKOJ]
25. R Foundation for Statistical Computing. 2017. R: A language and environment for statistical computing Internet URL: https://www.r-project.org/ [accessed 2018-04-05] [WebCite Cache ID 6yRUUNg1E]
26. World Health Organization. 2018. Family planning/Contraception URL: http://www.who.int/mediacentre/factsheets/fs351/en/ [WebCite Cache ID 6yRUdCZUV]
27. Kleiber C, Zeileis A. The Comprehensive R Archive Network. 2008. AER: Applied Econometrics with R URL: https://cran.r-project.org/web/packages/AER/index.html [accessed 2018-04-05] [WebCite Cache ID 6yRUiOW56]
28. The Comprehensive R Archive Network. 2014. ivpack: Instrumental Variable Estimation URL: https://cran.r-project.org/web/packages/ivpack/index.html [accessed 2018-04-05] [WebCite Cache ID 6yRUptSCz]
29. Angrist J, Pischke J. Mastering 'Metrics: The Path from Cause to Effect. Princeton, NJ: Princeton University Press; 2014.
30. Holland PW. Causal Inference, Path Analysis, and Recursive Structural Equations Models. Sociological Methodology 1988;18:449. [doi: 10.2307/271055]
31. Tatango. The Average SMS Marketing Click-Through Rate is 36% 2015 URL: https://www.tatango.com/blog/the-average-sms-marketing-click-through-rate-is-36/ [accessed 2018-04-05] [WebCite Cache ID 6yRV4Cbbq]
32. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive Health Behavior Change Text Message Interventions: A Meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
33. Smith C, Ngo TD, Gold J, Edwards P, Vannak U, Sokhey L, et al. Effect of a mobile phone-based intervention on post-abortion contraception: a randomized controlled trial in Cambodia. Bull World Health Organ 2015 Dec 01;93(12):842-50A [FREE Full text] [doi: 10.2471/BLT.15.160267] [Medline: 26668436]

Abbreviations

DHS: Demographic and Health Survey
LATE: local average treatment effect
SMS: short message service

Edited by G Eysenbach; submitted 11.04.18; peer-reviewed by R Grace, S Şahinöz; comments to author 16.05.18; accepted 14.06.18; published 31.07.18.

Please cite as:
Green EP, Augustine A, Naanyu V, Hess AK, Kiwinda L
Developing a Digital Marketplace for Family Planning: Pilot Randomized Encouragement Trial
J Med Internet Res 2018;20(7):e10756
URL: http://www.jmir.org/2018/7/e10756/
doi:10.2196/10756
PMID:30064968

Review

Barriers to and Facilitators of Engagement With Remote Measurement Technology for Managing Health: Systematic Review and Content Analysis of Findings

Sara Simblett[1], PhD, DClinPsy; Ben Greer[1], BSc; Faith Matcham[1], PhD, CPsychol; Hannah Curtis[1], BSc; Ashley Polhemus[2], MSE; José Ferrão[2], PhD; Peter Gamble[2], MBA; Til Wykes[1,3], DBE, PhD

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

[2]MSD IT Global Innovation Center, Prague, Czech Republic

[3]National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health, South London and Maudsley NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Sara Simblett, PhD, DClinPsy
Institute of Psychiatry, Psychology and Neuroscience
Psychology
King's College London
16 De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 2078480762
Email: sara.simblett@kcl.ac.uk

Abstract

Background: Remote measurement technology refers to the use of mobile health technology to track and measure change in health status in real time as part of a person's everyday life. With accurate measurement, remote measurement technology offers the opportunity to augment health care by providing personalized, precise, and preemptive interventions that support insight into patterns of health-related behavior and self-management. However, for successful implementation, users need to be engaged in its use.

Objective: Our objective was to systematically review the literature to update and extend the understanding of the key barriers to and facilitators of engagement with and use of remote measurement technology, to guide the development of future remote measurement technology resources.

Methods: We conducted a systematic review using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines involving original studies dating back to the last systematic review published in 2014. We included studies if they met the following entry criteria: population (people using remote measurement technology approaches to aid management of health), intervention (remote measurement technology system), comparison group (no comparison group specified), outcomes (qualitative or quantitative evaluation of the barriers to and facilitators of engagement with this system), and study design (randomized controlled trials, feasibility studies, and observational studies). We searched 5 databases (MEDLINE, IEEE Xplore, EMBASE, Web of Science, and the Cochrane Library) for articles published from January 2014 to May 2017. Articles were independently screened by 2 researchers. We extracted study characteristics and conducted a content analysis to define emerging themes to synthesize findings. Formal quality assessments were performed to address risk of bias.

Results: A total of 33 studies met inclusion criteria, employing quantitative, qualitative, or mixed-methods designs. Studies were conducted in 10 countries, included male and female participants, with ages ranging from 8 to 95 years, and included both active and passive remote monitoring systems for a diverse range of physical and mental health conditions. However, they were relatively short and had small sample sizes, and reporting of usage statistics was inconsistent. Acceptability of remote measurement technology according to the average percentage of time used (64%-86.5%) and dropout rates (0%-44%) was variable. The barriers and facilitators from the content analysis related to health status, perceived utility and value, motivation, convenience and accessibility, and usability.

Conclusions: The results of this review highlight gaps in the design of studies trialing remote measurement technology, including the use of quantitative assessment of usage and acceptability. Several processes that could facilitate engagement with this

technology have been identified and may drive the development of more person-focused remote measurement technology. However, these factors need further testing through carefully designed experimental studies.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42017060644; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=60644 (Archived by WebCite at http://www.webcitation.org/70K4mThTr)

(J Med Internet Res 2018;20(7):e10480) doi:10.2196/10480

KEYWORDS

mHealth; technology; engagement; systematic review; telemedicine; remote sensing technology; patient participation; review

Introduction

Global smartphone ownership has increased, which provides ready access to the internet, and a means of actively logging information and passively gathering big data [1]. Alongside this, a surge in the availability of wearable devices (eg, smart watches and fitness trackers) has enabled continuous and real-time collection of biosignatures and accelerometry [2]. These mobile tools, and platform infrastructures surrounding them, could provide intelligent remote measurement technology (RMT) to support health management. Direct feedback, for instance information about sleep quality, heart rate, mood, and activity, could enable users of RMT to play a more active role in managing their own health that is integrated into daily life. Similarly, feedback to health care professionals could facilitate efficient and timely decisions about treatment. Although these tools have the capacity to augment and extend health care opportunities, they also come with challenges associated with acceptability. A clear understanding of the key barriers to and facilitators of engagement for all stakeholders is an essential part of developing feasible, acceptable, and desired RMT systems.

Engagement is defined as the extent to and manner in which people actively use a resource and has been operationalized as a multistage process involving the point of engagement, a period of sustained engagement, disengagement, and reengagement [3]. Many factors may influence this engagement process at different time points. Indicators of poor engagement may include low initial uptake from the first point of contact or reduced interaction over time, in some cases leading to complete disengagement or dropout. Davis et al [4] conducted a systematic review of the feasibility and acceptability of RMT in primary care from the perspective of staff. They extracted themes from 16 studies, which included concerns regarding changes to roles and responsibilities, the need for extra resources and training, and questions about the usefulness of the data and overtreatment of patients. However, they also highlighted the benefits associated with direct patient education. They emphasized the need for target users, that is, people living with health problems, to be involved in product development and implementation, but the engagement of these target users was beyond the scope of their previous review.

The purpose of this systematic review was to update and extend the understanding of the barriers to and facilitators of engagement with RMT systems for target users. We defined RMT following Davis et al [4], and we categorized it into passive (data are obtained by on-body biosensors and built-in smartphone sensors) and active RMT (requires some interaction, such as completing short questionnaires at repeated time intervals). Passive RMT may interact with active RMT, by sensor activation prompts to perform an action. The review followed the population, intervention, comparison group, outcomes, and study design framework, to answer questions related to barriers to and facilitators of engagement with RMT systems. We achieved this through analysis of the qualitative feedback and quantitative data, such as ratings scales and usage statistics gathered from people using RMT. The aim was to extend the evidence in this area to guide the development of future RMT resources.

Methods

Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we conducted a systematic review of studies to answer the question "What are the barriers to and facilitators of engagement with remote measurement technology?" We registered the trial with the International Prospective Register of Systematic Reviews (PROSPERO registration number CRD42017060644).

Inclusion Criteria

We included studies if they met the following criteria: (1) were published in English; (2) included health care RMT, defined as any mobile technology that enables monitoring of a person's health status through a remote interface, with the data then either transmitted to a health care provider for review or to be used as a means of education for the user themselves [4]; and (3) were original studies published from January 2014 reporting the results of questionnaires, interviews, focus groups, and other indicators (eg, reasons for dropout), providing information about barriers to and facilitators of engagement with RMT systems using mHealth tools. We stipulated no diagnostic exclusions, so we included people using RMT to support any physical or mental health condition and healthy populations where interventions focused on improving general well-being.

Search Strategy

We searched Ovid MEDLINE, IEEE Xplore, EMBASE, Web of Science, and the Cochrane Library using the combined terms "remote" or "mobile" and "technology" or "devices," along with "telemedicine" and "mHealth." Multimedia Appendix 1 provides details of all search strategies. The initial search was completed in July 2016 and the process was repeated in May 2017. Two authors (SS and FM) independently screened articles by titles, abstracts, and then full texts to assess whether they

met the inclusion criteria. The repeated screening on the second batch of articles was carried out by 2 other authors (BG and HC).

Data Abstraction and Synthesis

Study Characteristics

We extracted the following data: (1) device type and RMT system (including active and passive data); (2) population characteristics, including diagnostic categories, sample size, time using RMT, and the country in which the study was conducted; and (3) methods used to gather qualitative information on the feasibility and acceptability, grouped as follows: usage statistics, questionnaires, structured or semistructured interviews, focus groups, and descriptive feedback.

Content Analysis

One author (SS) read and reread the results reported in articles published from January 2014 to July 2016 to extract individual barriers and facilitators (defined as "a circumstance or obstacle that may prevent the adoption of remote measurement technology" or "make adoption easy or easier"). The coding frame was developed by 3 authors (SS, BG, and HC) using these data. It consisted of the following themes: health status, usability, convenience and accessibility, perceived utility, and motivation, with subthemes. This coding frame was then tested on a further batch of articles published from June 2016 to May 2017 (coded by authors BG and HC and discrepancies evaluated by SS). This replication test allowed for a validation and potential extension of the initial coding frame.

Multimedia Appendix 2 and Multimedia Appendix 3 provide an overview of all coded barriers and facilitators. Some subthemes were mentioned as both a barrier and a facilitator depending on circumstances, and were coded separately. Multimedia Appendix 4 summarizes all quotes extracted and coded from each of the articles.

Assessing Study Quality

Methodological quality was assessed by 2 independent raters using the Mixed Methods Appraisal Tool (MMAT) [5]. The MMAT is a 21-item checklist of 5 research designs, with scores ranging from 0 to 1 in increments of 0.25. The MMAT does not provide a categorical distinction between studies of low or high quality; rather, it provides a descriptive framework of study quality. Interrater reliability has been reported to range from moderate to perfect (kappa range .53-1; Pace et al [5]).

Results

Study Selection

Of the 3187 abstracts and titles identified, 33 original articles met our inclusion criteria (see the PRISMA flow diagram in Figure 1 for a breakdown of this process). Multimedia Appendix 5 [6-38] presents study characteristics and participant demographics.

Participants

Studies varied in their sample size (7-365 participants), as well as the age (8-95 years) and sex of participants (30 studies included both male and female participants).

Study Characteristics

Studies were conducted in 10 countries: the United States (n=24), United Kingdom (n=1), Canada (n=1), Taiwan (n=1), Sweden (n=1), Poland (n=1), Australia (n=1), Switzerland (n=1), Germany (n=1), and New Zealand (n=1). Study durations ranged from 1 to 13 months, and 3 studies consisted of only a single individual or group session.

Remote Measurement Technology Characteristics

A total of 6 studies used passive RMT, including wearable pedometers and accelerometers, and built-in smartphone activity monitors (see Multimedia Appendix 5). Most studies used active RMT (n=17), including smartphone-based systems (eg, ecological momentary assessment, patient-reported outcome measures, and activity logs) and wireless monitoring devices (eg, blood pressure monitors and weight scales). Both active and passive RMT were used in 10 studies.

RMT systems provided feedback to users (n=17), members of the users' health care team (n=7), or both (n=9). Feedback was provided in various forms, including visual displays (eg, graphs), report summaries, historic reporting patterns, and messages (eg, health advice and motivational feedback).

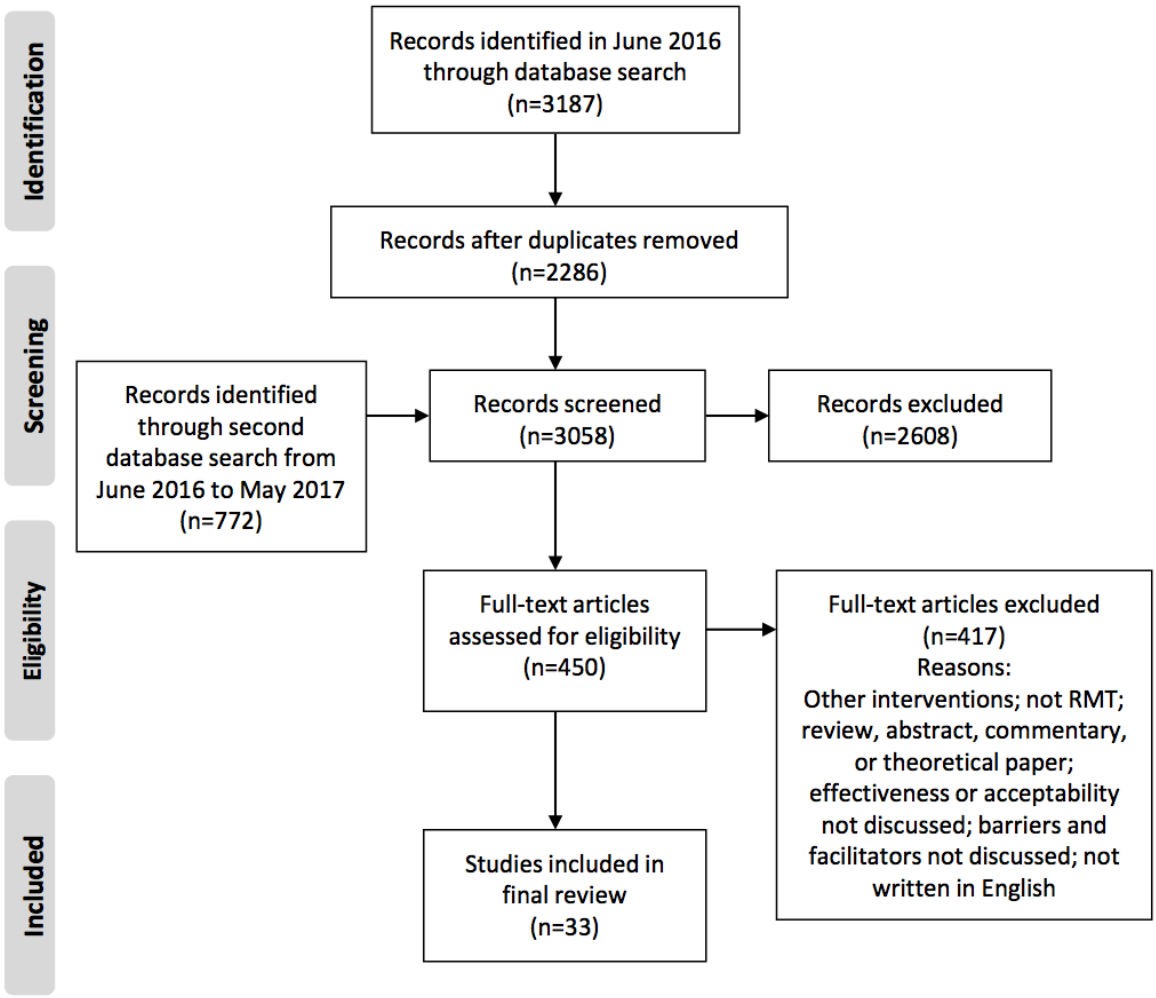
Health Conditions

The studies covered many health conditions, with most concentrating on 1 condition (n=17). A total of 2 studies featured more than 1 physical health diagnosis (diabetes and obesity, and multiple genetic blood disorders). Only 4 studies related to mental health conditions such as psychosis and posttraumatic stress disorder, and 2 studies included both physical and mental health conditions (eg, depression and type 2 diabetes, HIV, and substance use disorders). The remaining studies supported general health and well-being (n=7), and smoking cessation (n=1).

Assessment of Outcomes

In total, 27 studies employed quantitative methods to identify barriers to and facilitators of using RMT systems, including usage statistics (n=20) and questionnaires (n=19). Most questionnaires (15/19, 79%) were unvalidated measures developed for the study. Only 4 studies used validated measures, including the System Usability Scale, the Telehealth Usability Questionnaire, and the Technology Acceptance Model Questionnaire. Similarly, types of usage statistics reported varied greatly between studies. Of these 27 studies, 9 employed a mixed-methods design and asked for qualitative information (ie, from semistructured interviews and focus groups) and quantitative information from their users; 6 studies employed purely qualitative methods.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of study selection. RMT: remote measurement technology.



Study Quality

Of the reviewed studies, 2 obtained the maximum score of 1 on the MMAT [6,7], with the remaining studies scoring 0.75 (n=13), 0.5 (n=11), or 0.25 (n=7). Higher ratings were prohibited for a range of reasons, including a lack of adequately reported information regarding researchers' influence on the qualitative findings and their generalizability, description of sampling method, and method of analysis.

Quantitative Measures: Engagement and Adherence

Of the 5 studies that reported on the average number of times the RMT system was used, 3 reported the total number of interactions and 2 reported the number of days that people interacted with the app; 2 reported on the percentage of people who wore the wearable device for the whole study; and 4 set a threshold for the appropriate level of adherence (which varied between studies) and reported the percentage of people meeting these requirements. The remaining studies reported idiosyncratic usage statistics that were not comparable across studies. This variability severely limited quantified conclusions. For the few studies that reported the average percentage of time used, this ranged from 64% to 86.5% [8-10]. The average total number of interactions varied between 8.5 and 29.7 and may have depended on the type and length of the intervention [6,11,12]; the lowest level of interaction was with video content and the highest was with a person via a text message. The average numbers of interactions per week also varied between 3.5 [13] and 12 times per week [14]. The average percentage of people who wore the wearable device for the duration of the study ranged from 50% to 75% [15,16], and the percentage of people meeting a prespecified threshold for adherence varied from 41.7% to 81.8% [7,9,12,17]. Although studies reported varying degrees of attrition [10,13,18], dropout rate reporting was more frequent and ranged from 0% to 44% with a mean of 11.0% (SD 11.4). Table 1 summarizes the reasons reported for dropout. Overall, there was significant variation across studies, and there was no specific measure that is comparable across studies.

Table 1. Reasons for dropout across studies.

Reason for dropout	Frequency	Related theme
Lost or stolen smartphone	23	Usability
Technical malfunction (eg, smartphone corrupted, not receiving texts, or delivery delays)	7	Usability
Exacerbation of health condition, including participants who were injured or died during the course of the study	6	Health status
Deleted app	3	Usability
App not compatible with existing smartphone	3	Convenience and accessibility
Unexpected usage patterns (eg, switched smartphone off in between answering surveys, left smartphone plugged into charger, used smartphone in airplane mode)	3	Perceived utility
Moved out of area or was discharged from hospital	3	Convenience and accessibility
Sold smartphone	2	Perceived utility
Changed mobile phone or service plan	2	Convenience and accessibility
Practical technical difficulties (eg, not being able to download the app)	2	Usability
Broken smartphone	1	Usability
Inconsistent wireless network	1	Convenience and accessibility
App consumed too much battery	1	Usability
System too slow	1	Usability
Unspecified reason	11	Not applicable

Qualitative Analysis: Themes of Barriers and Facilitators

We divided themes into 5 major categories that made up a coding frame for structuring the minor themes. The two batches of articles (2014-2015 and 2016-2017) yielded subthemes that fitted within the same coding frame, with all major themes represented across the two time periods providing evidence of validity. No new themes arose in the later studies. The following section describes the findings for each major theme, with barriers and facilitators in italics. Multimedia Appendix 2 and Multimedia Appendix 3 display the categorization of subthemes for active RMT and passive RMT (including combinations of active and passive RMT), respectively.

Health Status

Exacerbations in health conditions, such as a chronic heart or respiratory condition, or episodes of being acutely unwell, such as experiencing a sickle cell crisis, have been reported to disrupt engagement and RMT use [6,10,18]. This disruption was related to a change in environment (hospital rather than own home) [6,18], as well as the acute exacerbation of health problems. Other longer-term health-related barriers to engagement in RMT included difficulties due to *poor vision* [19]. This was discussed in the context of older age; however, this was not tested directly.

Usability

Technical malfunctions were by far the most widely reported barriers, with 11 studies reporting ways in which these factors affected usability of the RMT systems [6,10,17,20-27]. This included not receiving notifications or receiving them at the wrong time, disappearance of the app, freezing of the system, losing power or restarting without warning, and difficulties connecting remote (wearable and other smart technology) devices with apps. Studies reported that this led to participant withdrawal [6], data loss [17,23,24], or significantly fewer data entries (eg, by 35%) [10].

Ben-Zeev et al [8] reported that *clarity of information* enhanced usability and facilitated engagement. In their study, 90% of participants reported that they thought they could learn to use the app very quickly, but no data were provided to suggest that these self-reports were valid. For other studies, difficulties inputting information into apps was a reason for discontinuing [15]. This may have depended on the type of data, length of time that participants were required to log data, or the value that people placed on the feedback, but a theme around engagement being potentially facilitated by clear and simple tasks emerged.

Where technical malfunctions and complexities in terms of usability arose, practical support was sometimes necessary. Some studies reported that problems such as “creating user accounts, answering intake question and navigating content due to unexpected behavior of keyboards, scroll bars, buttons, and other interface widgets” could be addressed with minor adjustments [22], although the authors provided no data on changes that had improved engagement. Engelhard et al [9] reported that where technical difficulties arose all could be solved by a phone call with the study coordinator; these authors offered no data to back up this claim.

In addition to technical functionality and clarity of information, we grouped other subthemes under the broader theme of usability. *Speed of the system* was a potential influence on engagement, with 1 participant withdrawing from a study due to frustrations with the slowness of the system [6]. *Use of larger devices* (smart tablets vs smartphones) in 1 study resulted in significantly more diary entries (by 30%) [10]. Given that this

difference emerged between 2 groups, it is unclear whether this arose from individual preferences or that larger devices led to better engagement. However, in another study that compared within-group differences, only 20% (10/51) of participants aged between 50 and 94 years were reported to be capable of using a smartphone, as opposed to a larger smart tablet for data entry and active monitoring; of these 10 people, only 3 considered the smaller device easy to use [28]. *Lost or damaged device* was a clear barrier to usability and participation, mentioned in 4 studies [8,13,15,26]. Further disruptions to response collection due to *changes in service plans* such that participants could no longer receive text messages [17] or excessive consumption of the *smartphone battery* [29] were mentioned as a barrier to data entry completion in another study [10].

Convenience and Accessibility

Compatibility with one's existing routine, including the ability to use your own devices, appeared as a subtheme. Ding et al [29] reported that 2 participants withdrew because the app was unable to function on their personal smartphone. Peng et al [30] stated that, even though the app functioned correctly, participants did not necessarily use it if other strategies, such as paper logbooks, already satisfied their needs. What is not clear from this study is at what point participants disengaged: immediately or after a trial period? Convenience was limited when there were *restrictions on the placement* of the wearable device—for example, participants had to carry their smartphone in their pocket [31]. However, resulting data loss was not reported. Systems that provided opportunities for *passive or automatic* data collection were endorsed as being more convenient where this approach met the objectives of RMT [32], but the impact on adherence was not a focus of the study.

Where users were required to actively engage with data collection (active RMT), the *presence of notifications* facilitated engagement [14,20]. These notifications became less important once the monitoring had become part of the participant's daily routine [14]. Surveys were much more likely to be completed if users were prompted with a notification. For instance, 93.5% of check-in surveys were completed following a notification rather than being self-initiated [12]. But other systems seemed to be able to produce high engagement even from self-initiated reports without prompts. For instance, a study by Ben-Zeev and colleagues achieved 62.5% adherence to data collected on mood, sleep, medication use, and psychosis symptoms [8]. But notifications can also be a barrier when they are not received at the right time; Cushing et al [20] and Juengst et al [23] and other studies reported that participants requested the ability to postpone responses to notifications so they might answer them at a convenient time [33], but there is no evidence that when this was done there was an improvement in engagement.

Other major barriers were related to participants' access to resources such as websites and videos due to a poor internet connection or lack of a Wi-Fi connection, and use of old computer systems [10,11,15]. This caused difficulties with specific processes such as setting up resources [21], with 2 participants withdrawing due to difficulties in acquiring a consistent wireless service [14]. Other problems with accessibility included *poor telephone network coverage*, which caused delays in receiving text messages [15] and, in 1 case, resulted in 39% of participants missing training sessions [7].

Lack of familiarity with and knowledge about how to use technology, such as websites, smartphone apps, and wearable devices, was reported as a challenge with using RMT systems and a source of frustration for participants [21,26]. But the impact on engagement was not quantified. *Forgetfulness* was raised as interfering with the individual's ability to access passwords, complete questionnaires, wear their device, and sync their wearable device to their smartphone [10,16,21,26], but this was not quantified. Digital literacy and other practical barriers were overcome through offering instructions and *support* from the study coordinator [9,28]. Research into the type of support necessary to increase engagement was lacking and may be a subject for future reviews.

Other barriers within this theme included RMT systems not being adequately *tailored* to the disability status of individual participants. In the study of Engelhard et al [9], some participants felt that questions were irrelevant to them and did not want to continue reporting symptoms that showed no sign of change. The authors suggested integrating adaptive patient-reported outcome measures. Cultural relevance of study support materials was also reported to enhance engagement [25]; however, this was a qualitative study that provided no evidence of how it enhanced engagement.

Perceived Utility

Perceived Rewards

The results of 4 studies demonstrated a positive and motivating effect of feedback [11,32,34,35]. Buchem et al [34] reported that 50% of participants felt motivated by virtual rewards such as badges (ie, an indicator of accomplishment, skill, quality, or interest that can be earned). Dale et al [11] reported that 67% of participants liked receiving motivational texts from the RMT system. The results were less clear in the remaining studies, but some participants reported a benefit associated with learning about their real-time activity [32] and talking about app data with a study coordinator [35].

Further *incentives* that were suggested to increase motivation to engage included social sharing and comparison [16,32,36] or gaming features, including monetary rewards [20]. Another aspect reported to be "enjoyable" in 1 study was the receiving the training instructions, which was seen to be an important contributor toward increased engagement [34].

Perceived Costs

Financial costs were a clear barrier to engagement in 2 studies. Ho et al [36] found that 56% of their sample, based on their current income, would have struggled to afford a program that required payment of a large initial sum, followed by smaller regular payments. Naslund et al [16] reported that commercially available, wearable tracking devices alone were seen to be expensive and difficult to obtain for individuals with a low income. Some participants who were provided with devices that were perceived to be expensive were found to sell or pawn them [13].

Privacy concerns were also reported in 1 study, in which a participant disengaged and switched their mobile phone to airplane mode due to concerns about being tracked [13]. This study investigated an RMT system for people with psychosis and was the only study to raise concerns about privacy as the reason for disengagement. Disengagement was, however, raised in relation to other issues such as feeling uncertain about the user benefits and the *reliability or accuracy* of the data being recorded [15,28].

Motivation

The value of the RMT system appeared to be affected by people's *intrinsic motivation* to learn and sustain engagement. The impact of perceived rewards on motivation has already been mentioned, but these studies did not quantify this effect or report the impact across time. One additional study highlighted that, over time, active RMT became burdensome, and this affected 1 participant's motivation to engage [30]. Others reported that boredom had a negative impact on engagement [32]. The magnitude of this negative impact was not measured and discussed. *Extrinsic motivation* and reception from others (eg, clinicians) also affected use, with participants reporting a reluctance to try mHealth technologies if their doctor did not recommend it [30]. However, this finding was reported in the context of a hypothetical scenario rather than in a trial of an actual RMT.

Relationship Between Adherence and Themes

Dropout is a clear indicator of problems with engagement. Reasons for dropout spanned several of the qualitative themes, with problems related to usability of the wearable device and the smartphones apps being the most frequent. Convenience and accessibility was the second most frequent theme. The study that reported the greatest percentage of dropouts included one of the largest samples (n=342) and followed people with a diagnosis of psychosis for 6 months. Studies that reported no dropouts or the odd person dropping out were much smaller (ranging from 8 to 51 participants), and dropout may not be possible to understand here, as the sample might have been highly selective. There was no significant relationship between the percentage of people who dropped out and the length of the intervention in days (r_{29}=.19, P=.31).

A total of 10 studies reported on the impact of variables on adherence in terms of compliance and use of an mHealth device over time. The themes included health status, with greater physical disability [9] and mental health problems (symptoms of posttraumatic stress disorder) [17] being associated with better engagement (ie, participants exceeded usage requirements and provided more responses, respectively), but rehospitalization being a barrier to engagement [13,18]. Issues to do with usability was the second most common category, with technical difficulties accounting for poorer compliance (eg, missed assessments) [10,23], and use of larger mobile tablet, as compared with a smartphone, being significantly higher [10,28]. Confidence in one's ability to maintain an exercise regimen correlated with percentage of ecological momentary assessment responses [33]. Sociodemographic factors have also been found to influence use of mHealth technology, with age appearing to moderate use [10,28]. Lower household income, higher level of education, and male sex have been found to be facilitators for mHealth technology use [9,28,34].

Discussion

Factors Driving Engagement

Many of the factors discovered are consistent with the engagement attributes previously reported by O'Brien and Toms [3] in their model for engagement with technology. They described a dynamic model, where engagement is a continual cycle of engagement, disengagement, and reengagement that persists over time. While they described many factors that drive engagement with technology in general, RMT to manage health outcomes is a specific and unique technology, in which health-related symptoms and potential moderators offered by health care providers should be considered. Building on this work and using themes from this review, we present a model of the most prominent influences on RMT engagement, including key facilitators (Figure 2).

Engagement in our model is moderated by health status, usability, convenience and accessibility, perceived utility, and motivation to engage. Engagement may be at its strongest when the user is able to use the technology, perceives the technology to be useful, and wants to use the technology.

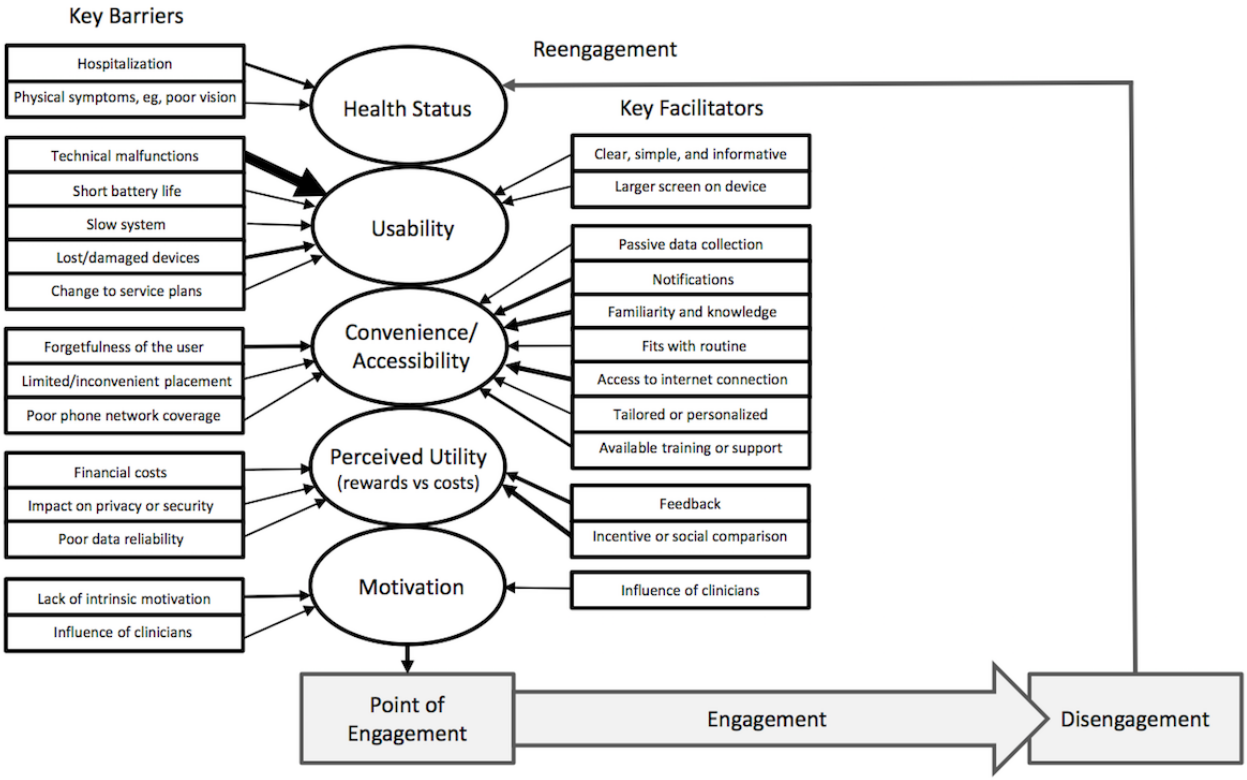
Health Status

Of particular importance to RMT systems for management of health outcomes is the health status of the user. Health status will inevitably have an impact on what constitutes a usable, convenient, accessible, or valuable feature of an RMT system. As an example, being unwell and outside of one's usual environment or routine (eg, in the hospital) led to disruptions in engagement and dropout [6,10,18]. However, some evidence suggests that people who were experiencing a higher level of problems (eg, greater physical or mental disability) engaged better [9,17]. While health severity and need for support may increase one's motivation to participate, factors such as health condition and disability status, including typical or fluctuating symptoms, should always be considered in the design and implementation of RMT systems for management of health outcomes.

Usability

At the heart of this proposed model is usability. There may be individual differences that moderate usability, including variables such as age, past experience with technology, and exacerbations in health conditions and disability status, as well as the influence of how the system is designed. Problems with usability were the most common reasons for dropout from the studies. There is evidence that older adults were harder to engage [19,28,37]. This was partly because some were unfamiliar with using mHealth tools such as smartphone and wearable devices or did not feel motivated to learn new skills, but also because the devices were of unsuitable size to accommodate changing needs (eg, larger, more legible font sizes). Where content is presented clearly, such as in a smartphone app, and adequate support (actions or resources designed to help users work through challenges posed by the system) is offered, engagement seems to be facilitated [8].

Figure 2. Model of barriers to and facilitators of engagement with remote measurement technology.



However, the specific parameters for this support are unknown and need further research with clearly quantifiable outcomes. In addition, involvement of user experience methods is important for the development of usable mHealth tools for RMT systems in the future, with coproduction and user-centered design processes to validate choices [39].

Convenience and Accessibility

The need to be able to integrate the RMT system into a user's normal routine was clear. Participants preferred tools that fit in with daily routines and tools that have already been adopted, with some disengaging and dropping out if unacceptable alternatives were offered. Personalization and demonstrating flexibility, in terms of taking into account the specific disabilities and needs of clinical groups, may be key in the design of usable RMT systems. This may include individual goal setting of dates and times for study activities, opting in or out of certain tracking activities (eg, reducing intrusiveness), or accommodating for health-related differing abilities. It may be important to note that forgetfulness emerged as a key barrier to engagement, which may suggest that the cognitive burden placed on individuals to remember to complete RMT schedules, in these studies, was too great. The value of notifications and reminders to carry out tasks has been demonstrated through usage statistics. That said, the magnitude of the effect varied between studies, with 1 study demonstrating a much bigger impact of notifications. This suggests that other factors moderate the likelihood of self-initiated engagement. Prompts have been mentioned to help aid memory, but there was some suggestion that the timing of these strategies may be important [17] and that there may be individual differences in preferences, with notifications that are too frequent being experienced as intrusive [32], thereby increasing cognitive burden. However, the studies did not manipulate these factors in an experimental design to test their impact, and this needs further research. Additional practical problems, such as poor Wi-Fi access, mobile data and network coverage, or compatible devices proved prohibitive to engagement [14,15,21]. Individual adaptation is exemplified by the size of devices. In some studies participants wanted smaller, more portable devices [31], and in other studies participants expressed the desire to have bigger monitors to be able to see their health data and complete the surveys more easily [19]. Balancing these goals may be a challenge for the development of future resources and may require coproduction with users to determine what is acceptable given a specific context. Some flexibility may be possible, for example, the use of responsive app designs that scale to the device being used. However, with the likelihood of large individual variation, this will be a major challenge for implementing RMT. Further research is needed to better quantify the magnitude of other potential facilitators that may help to overcome the barriers associated with convenience and accessibility.

Perceived Utility

We propose that increasing the rewards of using RMT increases the overall perceived value of the system in the face of some potential costs. Costs included financial costs of purchasing equipment, as well as concerns about privacy and reliability or accuracy of the data collected. As a strategy for increasing rewards associated with RMT systems, feedback is generally accepted, tolerated, and, in some cases, actively sought by users of RMT systems. In this context, feedback is considered to be additional information that participants receive from an RMT system about their health, their participation, or the larger

program from which users and participants can derive value. This could include health information, rates of participation or adherence, metrics defined in goal-setting exercises, positive reinforcements, or general information about the study or their health condition. It was commonly reported that participants would like to receive more feedback [17,22,26], with some concluding that future efforts to improve long-term engagement should include positive reinforcements [10]. There is some emerging evidence for a role of social comparison and of incentives through gamified competition and monetary rewards on maintaining engagement. What is not yet known is what is the most effective method of providing feedback and incentives, and it may be important to note that perceptions of reward may differ between individuals. People with more severe health problems may be more likely to engage with RMT. This may be linked to perceived utility, as people with worse health status may perceive greater potential benefit to using the RMT.

Motivation

Motivation was a smaller but important category emerging from the analysis of the results of previous studies using RMT systems for the management of health outcomes. Without motivation, participants may not engage with the initial process of learning how to use a new system, and this category is inextricably linked to all other factors discussed previously. Even if users are familiar with mHealth tools such as smartphones and wearable devices, they may need additional motivation to integrate a new set of behaviors, such as responding to surveys. Lack of motivation is therefore a fundamental barrier to engagement. The factors presented thus far should be considered not just at the initiation of the study, but also as engagement is managed over time, because perceptions of the technology's value or usability may change with prolonged use (eg, if expectations are not met). Therefore, we recommend steps to increase, or mitigate decreased, motivation with an RMT system to maintain motivation, and therefore engagement, over time.

Limitations of Previous Research and Future Recommendations

Facilitators identified include convenience and accessibility, perceived utility, and motivation, but these factors are drawn from of pool of studies that varied greatly in terms of their quality. In addition, we conceptualized engagement as a process that should include disengagement and reengagement when required, but most findings reported in the studies included in this review relate to moderators of initial and sustained engagement. Although in our model we tentatively propose a feedback loop between the point of disengagement and the same barriers and facilitators affecting initial and sustained engagement, it is possible that factors affecting reengagement may be different, and this was not the focus of the studies. Future research should focus on the entire engagement process and quantify the impact of specific variables on engagement in terms of observable changes in usage statistics in rigorous experimental design. Some examples might be looking at the impact of different types of support (automated messages vs personalized messages vs direct human support) on the number of interactions and overall time spent using a smartphone app or wearable device. The impact of different types of feedback (immediate vs delayed vs no feedback) and data visualization or communication methods (graphs vs text messages vs discussion with a study coordinator) or environment (hospital vs home-based use) also need to be explicitly tested. Careful experimental manipulation is missing from the literature to date and, to be able to compare across these conditions, quantitative measures and usage statistics also require more standardization. A similar conclusion has also been drawn when considering adherence [40]. As a minimum, the number of interactions with apps (both total interactions and numbers of days) and time spent wearing devices relative to the length of the trial needs to be collected.

It is not enough for software developers to consider their systems in isolation from the individuals who may be using them. One of the main ways to develop engaged systems is to begin with codesign with those individuals who will be using the system. This is especially important for those involved in providing RMT for improving health. Before RMT systems are tested, there needs to be an iterative design process that explores acceptability, such as following the principles of user-centered design [41,42]. The feedback gathered may be qualitative, and some of this exploratory work has been conducted and forms the basis of the model we present in this paper. However, this work needs to lead into quantitative assessment as described above.

Conclusions

The themes discovered in this review emerged across two different time periods providing validity information, but this evidence suggests that we are continuing to make the same mistakes. There is a great potential for RMT systems to augment and extend health care, but there remain clear challenges that still need to be overcome. Two suggestions are, first, to improve how we measure the impact of modifiable variables on engagement in order to understand the magnitude of effects. Second, several studies suggest working with the target users directly to coproduce systems that are acceptable and feasible to use over long periods of time. Our model indicates the interrelationship between key facilitators on the one hand, and the person and RMT factors on the other, that could act as a prototype for the development of RMT in the future.

Acknowledgments

This paper was written as part of the development of useful mHealth and remote measurement technology systems in the Remote Assessment of Disease and Relapse – Central Nervous System (RADAR-CNS) project. We acknowledge all partners in the RADAR-CNS consortium (www.radar-cns.org) for overall discussion of the results. The RADAR-CNS project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and European Federation of

Pharmaceutical Industries and Associations (EFPIA; www.imi.europa.eu). This communication reflects the views of the RADAR-CNS consortium and neither the Innovative Medicines Initiative nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. This paper also presents independent research funded in part by the UK National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the UK Department of Health and Social Care. TW would also like to acknowledge support from the NIHR Biomedical Research Centre at the South London and Maudsley Foundation Trust and King's College London, as well as the NIHR Senior Investigator Awards.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy

[PDF File (Adobe PDF File), 18KB - jmir_v20i7e10480_app1.pdf]

Multimedia Appendix 2

Facilitators of and barriers to engagement in active RMT.

[PDF File (Adobe PDF File), 70KB - jmir_v20i7e10480_app2.pdf]

Multimedia Appendix 3

Facilitators of and barriers to engagement in passive and combination RMT.

[PDF File (Adobe PDF File), 62KB - jmir_v20i7e10480_app3.pdf]

Multimedia Appendix 4

Systematic review quotes

[PDF File (Adobe PDF File), 57KB - jmir_v20i7e10480_app4.pdf]

Multimedia Appendix 5

Characteristics of the original studies included in the systematic review.

[PDF File (Adobe PDF File), 57KB - jmir_v20i7e10480_app5.pdf]

References

1. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies but advanced economies still have higher rates of technology use. Washington, DC: Pew Research Center; 2016 Feb 22. URL: http://assets.pewresearch.org/wp-content/uploads/sites/2/2016/02/pew_research_center_global_technology_report_final_february_22__2016.pdf [accessed 2018-06-19] [WebCite Cache ID 70IdHrDtG]
2. Morton N, Blackman D. The Growing Availability of Wearable Devices: A Perspective on Current Applications in Clinical Trials. Iselin, NJ: Applied Clinical Trials; 2016.
3. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
4. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]
5. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. Int J Nurs Stud 2012 Jan;49(1):47-53. [doi: 10.1016/j.ijnurstu.2011.07.002] [Medline: 21835406]
6. Hardinge M, Rutter H, Velardo C, Shah SA, Williams V, Tarassenko L, et al. Using a mobile health application to support self-management in chronic obstructive pulmonary disease: a six-month cohort study. BMC Med Inform Decis Mak 2015 Jun 18;15:46 [FREE Full text] [doi: 10.1186/s12911-015-0171-5] [Medline: 26084626]

7. Piotrowicz E, Korzeniowska-Kubacka I, Chrapowicka A, Wolszakiewicz J, Dobraszkiewicz-Wasilewska B, Batogowski M, et al. Feasibility of home-based cardiac telerehabilitation: results of TeleInterMed study. Cardiol J 2014;21(5):539-546. [doi: 10.5603/CJ.a2014.0005] [Medline: 24526507]
8. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253. [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
9. Engelhard MM, Patek SD, Sheridan K, Lach JC, Goldman MD. Remotely engaged: Lessons from remote monitoring in multiple sclerosis. Int J Med Inform 2017 Dec;100:26-31. [doi: 10.1016/j.ijmedinf.2017.01.006] [Medline: 28241935]
10. Jonassaint CR, Shah N, Jonassaint J, De CL. Usability and feasibility of an mHealth intervention for monitoring and managing pain symptoms in sickle cell disease: the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2015;39(3):162-168. [doi: 10.3109/03630269.2015.1025141] [Medline: 25831427]
11. Pfaeffli DL, Whittaker R, Dixon R, Stewart R, Jiang Y, Carter K, et al. Acceptability of a mobile health exercise-based cardiac rehabilitation intervention: a randomized trial. J Cardiopulm Rehabil Prev 2015;35(5):312-319. [doi: 10.1097/HCR.0000000000000125] [Medline: 26181037]
12. McClure JB, Anderson ML, Bradley K, An LC, Catz SL. Evaluating an adaptive and interactive mHealth smoking cessation and medication adherence program: a randomized pilot feasibility study. JMIR Mhealth Uhealth 2016 Aug 03;4(3):e94 [FREE Full text] [doi: 10.2196/mhealth.6002] [Medline: 27489247]
13. Ben-Zeev D, Scherer EA, Gottlieb JD, Rotondi AJ, Brunette MF, Achtyes ED, et al. mHealth for schizophrenia: patient engagement with a mobile phone intervention following hospital discharge. JMIR Ment Health 2016;3(3):e34 [FREE Full text] [doi: 10.2196/mental.6348] [Medline: 27465803]
14. Dicianno BE, Fairman AD, McCue M, Parmanto B, Yih E, McCoy A, et al. Feasibility of using mobile health to promote self-management in spina bifida. Am J Phys Med Rehabil 2016 Jun;95(6):425-437. [doi: 10.1097/PHM.0000000000000400] [Medline: 26488144]
15. Eisenhauer CM, Hageman PA, Rowland S, Becker BJ, Barnason SA, Pullen CH. Acceptability of mHealth technology for self-monitoring eating and activity among rural men. Public Health Nurs 2017 Mar;34(2):138-146. [doi: 10.1111/phn.12297] [Medline: 27757986]
16. Naslund JA, Aschbrenner KA, Barre LK, Bartels SJ. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. Telemed J E Health 2015 Mar;21(3):213-216. [doi: 10.1089/tmj.2014.0105] [Medline: 25536190]
17. Price M, Ruggiero KJ, Ferguson PL, Patel SK, Treiber F, Couillard D, et al. A feasibility pilot study on the use of text messages to track PTSD symptoms after a traumatic injury. Gen Hosp Psychiatry 2014;36(3):249-254 [FREE Full text] [doi: 10.1016/j.genhosppsych.2014.02.004] [Medline: 24636721]
18. Aranki D, Kurillo G, Yan P, Liebovitz DM, Bajcsy R. Real-time tele-monitoring of patients with chronic heart-failure using a smartphone: lessons learned. IEEE Trans Affect Comput 2016 Jul 1;7(3):206-219. [doi: 10.1109/TAFFC.2016.2554118]
19. Evangelista LS, Moser DK, Lee J, Moore AA, Ghasemzadeh H, Sarrafzadeh M, et al. Examining older adults' perceptions of usability and acceptability of remote monitoring systems to manage chronic heart failure. Gerontol Geriatr Med 2015 Nov;1:2333721415618050 [FREE Full text] [doi: 10.1177/2333721415618050] [Medline: 28138479]
20. Cushing A, Manice MP, Ting A, Parides MK. Feasibility of a novel mHealth management system to capture and improve medication adherence among adolescents with asthma. Patient Prefer Adherence 2016;10:2271-2275 [FREE Full text] [doi: 10.2147/PPA.S115713] [Medline: 27853357]
21. Fontil V, McDermott K, Tieu L, Rios C, Gibson E, Sweet CC, et al. Adaptation and feasibility study of a digital health program to prevent diabetes among low-income patients: results from a partnership between a digital health company and an academic research team. J Diabetes Res 2016;2016:8472391 [FREE Full text] [doi: 10.1155/2016/8472391] [Medline: 27868070]
22. Hartzler A, Venkatakrishnan A, Mohan S, Silva M, Lozano P, Ralston J, et al. Acceptability of a team-based mobile health (mHealth) application for lifestyle self-management in individuals with chronic illnesses. 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 17-20, 2016; Lake Buena Vista, FL, USA p. 3277-3281.
23. Juengst SB, Graham KM, Pulantara IW, McCue M, Whyte EM, Dicianno BE, et al. Pilot feasibility of an mHealth system for conducting ecological momentary assessment of mood-related symptoms following traumatic brain injury. Brain Inj 2015 Aug;29(11):1351-1361. [doi: 10.3109/02699052.2015.1045031] [Medline: 26287756]
24. Leonard S, Anderson LM, Jonassaint J, Jonassaint C, Shah N. Utilizing a novel mobile health "selfie" application to improve compliance to iron chelation in pediatric patients receiving chronic transfusions. J Pediatr Hematol Oncol 2017 Apr;39(3):223-229. [doi: 10.1097/MPH.0000000000000743] [Medline: 28099398]
25. Maglalang DD, Yoo GJ, Ursua RA, Villanueva C, Chesla CA, Bender MS. "I don't have to explain, people understand": acceptability and cultural relevance of a mobile health lifestyle intervention for Filipinos with type 2 diabetes. Ethn Dis 2017;27(2):143-154. [doi: 10.18865/ed.27.2.143] [Medline: 28439185]
26. Randriambelonoro M, Chen Y, Pu P. Can fitness trackers help diabetic and obese users make and sustain lifestyle changes? Computer 2017 Mar;50(3):20-29. [doi: 10.1109/MC.2017.92]

27. Spring B, Pellegrini CA, Pfammatter A, Duncan JM, Pictor A, McFadden HG, et al. Effects of an abbreviated obesity intervention supported by mobile technology: the ENGAGED randomized clinical trial. Obesity (Silver Spring) 2017 Jul;25(7):1191-1198 [FREE Full text] [doi: 10.1002/oby.21842] [Medline: 28494136]
28. Huang F, Chang P, Hou I, Tu M, Lan C. Use of a mobile device by nursing home residents for long-term care comprehensive geriatric self-assessment: a feasibility study. Comput Inform Nurs 2015 Jan;33(1):28-36. [doi: 10.1097/CIN.0000000000000115] [Medline: 25397723]
29. Ding X, Xu J, Wang H, Chen G, Thind H, Zhang Y. WalkMore: promoting walking with just-in-time context-aware prompts. 2016 Presented at: Wireless Health 2016; Oct 25-27, 2016; Bethesda, MD, USA p. 65-72.
30. Peng W, Yuan S, Holtz BE. Exploring the challenges and opportunities of health mobile apps for individuals with type 2 diabetes living in rural communities. Telemed J E Health 2016 Sep;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]
31. Al Ayubi SU, Parmanto B, Branch R, Ding D. A persuasive and social mhealth application for physical activity: a usability and feasibility study. JMIR Mhealth Uhealth 2014;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
32. 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]
33. Mundi MS, Lorentz PA, Grothe K, Kellogg TA, Collazo-Clavell ML. Feasibility of smartphone-based education modules and ecological momentary assessment/intervention in pre-bariatric surgery patients. Obes Surg 2015 Oct;25(10):1875-1881. [doi: 10.1007/s11695-015-1617-7] [Medline: 25702141]
34. Buchem I, Merceron A, Kreutel J, Haesner M, Steinert A. Gamification designs in wearable enhanced learning for healthy ageing. 2015 Presented at: International Conference on Interactive Mobile Communication Technologies and Learning; Nov 19-20, 2015; Thessaloniki, Greece p. 9-15. [doi: 10.1109/IMCTL.2015.7359545]
35. Vathsangam H, Sukhatme G. Using phone-based activity monitors to promote physical activity in older adults: a pilot study. 2014 Presented at: 2014 IEEE Healthcare Innovation Conference; Oct 8-10, 2014; Seattle, WA, USA p. 42-47. [doi: 10.1109/HIC.2014.7038870]
36. Ho K, Newton L, Boothe A, Novak-Lauscher H. Mobile Digital Access to a Web-enhanced Network (mDAWN): assessing the feasibility of mobile health tools for self-management of type-2 diabetes. AMIA Annu Symp Proc 2015;2015:621-629 [FREE Full text] [Medline: 26958197]
37. Westergaard RP, Genz A, Panico K, Surkan PJ, Keruly J, Hutton HE, et al. Acceptability of a mobile health intervention to enhance HIV care coordination for patients with substance use disorders. Addict Sci Clin Pract 2017 Dec 26;12(1):11 [FREE Full text] [doi: 10.1186/s13722-017-0076-y] [Medline: 28441962]
38. Lind L, Carlgren G, Karlsson D. Old-and with severe heart failure: telemonitoring by using digital pen technology in specialized homecare: system description, implementation, and early results. Comput Inform Nurs 2016 Aug;34(8):360-368. [doi: 10.1097/CIN.0000000000000252] [Medline: 27223309]
39. 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;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
40. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
41. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Volume 37 number 4. Thousand Oaks, CA: Sage Publications; 2004:445-456.
42. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]

Abbreviations

MMAT: Mixed Methods Appraisal Tool
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RMT: remote measurement technology

Edited by G Eysenbach; submitted 23.03.18; peer-reviewed by J Rawstorn, S Rush; comments to author 26.04.18; revised version received 09.05.18; accepted 10.05.18; published 12.07.18.

Please cite as:
Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, Gamble P, Wykes T
Barriers to and Facilitators of Engagement With Remote Measurement Technology for Managing Health: Systematic Review and Content Analysis of Findings
J Med Internet Res 2018;20(7):e10480
URL: http://www.jmir.org/2018/7/e10480/
doi:10.2196/10480
PMID:30001997

Original Paper

Predicting Mood Disturbance Severity with Mobile Phone Keystroke Metadata: A BiAffect Digital Phenotyping Study

John Zulueta[1], MD; Andrea Piscitello[1], MS; Mladen Rasic[1], BA; Rebecca Easter[1], BA; Pallavi Babu[2], BA; Scott A Langenecker[1], PhD; Melvin McInnis[2], MD; Olusola Ajilore[1], MD, PhD; Peter C Nelson[1], PhD; Kelly Ryan[2], PhD; Alex Leow[1], MD, PhD

[1]University of Illinois at Chicago, Chicago, IL, United States

[2]University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Alex Leow, MD, PhD
University of Illinois at Chicago
1601 W Taylor St
Chicago, IL, 60612
United States
Phone: 1 312 996 7383
Fax: 1 312 996 9517
Email: aleow@psych.uic.edu

Abstract

Background: Mood disorders are common and associated with significant morbidity and mortality. Better tools are needed for their diagnosis and treatment. Deeper phenotypic understanding of these disorders is integral to the development of such tools. This study is the first effort to use passively collected mobile phone keyboard activity to build deep digital phenotypes of depression and mania.

Objective: The objective of our study was to investigate the relationship between mobile phone keyboard activity and mood disturbance in subjects with bipolar disorders and to demonstrate the feasibility of using passively collected mobile phone keyboard metadata features to predict manic and depressive signs and symptoms as measured via clinician-administered rating scales.

Methods: Using a within-subject design of 8 weeks, subjects were provided a mobile phone loaded with a customized keyboard that passively collected keystroke metadata. Subjects were administered the Hamilton Depression Rating Scale (HDRS) and Young Mania Rating Scale (YMRS) weekly. Linear mixed-effects models were created to predict HDRS and YMRS scores. The total number of keystrokes was 626,641, with a weekly average of 9791 (7861), and that of accelerometer readings was 6,660,890, with a weekly average 104,076 (68,912).

Results: A statistically significant mixed-effects regression model for the prediction of HDRS-17 item scores was created: conditional R^2=.63, P=.01. A mixed-effects regression model for YMRS scores showed the variance accounted for by random effect was zero, and so an ordinary least squares linear regression model was created: R^2=.34, P=.001. Multiple significant variables were demonstrated for each measure.

Conclusions: Mood states in bipolar disorder appear to correlate with specific changes in mobile phone usage. The creation of these models provides evidence for the feasibility of using passively collected keyboard metadata to detect and monitor mood disturbances.

(J Med Internet Res 2018;20(7):e241) doi:10.2196/jmir.9775

KEYWORDS

digital phenotype; mHealth; ecological momentary assessment; keystroke dynamics; bipolar disorder; depression; mania; mobile phone

Introduction

The burden of mental illness is high. It has been estimated that mental illness accounts for 32% of years lived with disability around the world [1]. Bipolar disorder is a serious mental illness characterized by recurrent episodes of depression and mood elevation [2] and is associated with high rates of functional impairment, decreased quality of life, and increased rates of mortality from comorbid medical conditions [3]. Given these costs, it is imperative that we deepen our understanding of this disorder to promote accurate diagnosis and effective treatment.

The ubiquity of mobile phones, smartphones in particular, presents a new opportunity in the study of mental illness. An estimated 64% of adults in the United States own a mobile phone and use it for a variety of tasks, including phone calls, Web browsing, and social media; however, the most widely and frequently used feature on mobile phones is short message service text messaging [4]. These devices can be employed as platforms for the unobtrusive collection of myriad data that can be used in the study of psychopathology. Ecological momentary assessment is a methodology that aims to collect data using repeated measures in real time (or near real time), in people's natural environment [5]. When applied to the use of digital technologies such as mobile phones, this methodology can be used to create digital phenotypes defined as the set of observable behaviors resulting from the interaction between human disease and people's use of digital technologies [6].

Because recurring mood episodes are a defining characteristic of bipolar disorder, we posited that it is an ideal illness for a pilot study investigating the relationship between mobile phone keyboard activity and the correlates of these episodes, such as changes in cognitive function, psychomotor activity, social behavior, and diurnal activity patterns. We elected to focus on keystroke dynamics because features using text input (eg, texting and Web browsing) are among the most commonly used features in mobile phones and because we hypothesized that keystroke dynamics provide a sufficiently dense space from which to extract relevant features that could be used to predict the severity of depression and mania.

Methods

Participants

Study subjects were members of the Prechter Longitudinal Study of Bipolar Disorder, a naturalistic, longitudinal study based in the University of Michigan [7]. This cohort includes subjects with bipolar disorder, other psychiatric illnesses, and healthy controls; however, only those with bipolar disorder were recruited into this study. Subjects were recruited into this study by email or phone invitation. The inclusion criteria included being a current Android mobile phone user, asserting familiarity with the Android operating system, having no gross impairments in fine motor abilities, sufficient vision to use a mobile phone keyboard, and self-reporting of frequent mood fluctuations or having longitudinal data from the longitudinal study suggesting that they experience frequent mood symptoms (ie, endorsed frequent mood symptoms on bimonthly self-report measures of mood or categorized as rapid cycling).

We initially included 19 subjects with a bipolar spectrum disorder as per the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (Text Revision) criteria [8] (11 with bipolar I, 7 with bipolar II, and 1 with bipolar not otherwise specified); of these subjects, 1 never activated the app and 2 deleted the app early in the study. Of the remaining 16 subjects, participation varied in terms of the number of weeks that had any keyboard activity, with an average of 4.69 (3.05) weeks. Because of concerns about adherence, data analysis was restricted to subjects who provided data for at least 4 weeks. This resulted in 9 subjects: 5 with bipolar I and 4 with bipolar II. Of these, 8 subjects met the criteria for rapid cycling (ie, 4 or more mood episodes per year), and all subjects with bipolar II had recurrent depressive episodes. Of these 9 subjects, 7 showed keyboard activity for at least 6 weeks. The total usable data from these subjects included 626,641 keystrokes and 6,660,890 accelerometer readings.

Mobile Keyboard

A custom keyboard called "BiAffect" was developed for the Android operating system that replaced the default keyboard and collected metadata consisting of keystroke entry date and time and accelerometer displacement. It uploaded these data using secure encrypted protocols to the study server hosted at the University of Illinois at Chicago. Accelerometer data collection was initiated by keystroke entry and continued for 5 seconds afterward. Individual character data outside of the backspace key and space bar were not collected, anonymizing the entry. The keyboard was designed to appear similar to the standard Android keyboard (Figure 1).

Data Collection

For 8 weeks, subjects were provided a Samsung Galaxy Note 4 smartphone that they were instructed to use as their primary phone during the study period. Subjects were encouraged to use their current phone number and subscriber identification module card; with the exception of 1 subject, all subjects did so. During the study period, trained staff at the University of Michigan administered the Structure Interview Guide for the Hamilton Depression Rating Scale (HDRS) [9] and Young Mania Rating Scale (YMRS) [10] once a week via phone interviews.

Statistical Analyses

Subject demographics are described in Table 1. The YMRS results showed a right-tailed skew (γ_1=1.14) [11], so a log transformation was performed on the YMRS scores by taking the natural log of the sum of the YMRS scores and 1 (γ_1=−0.44).

In order to identify the possible relationships between subject demographics and phone usage, Spearman correlations were calculated between subjects' total key counts and their age and education.

Mixed-effects linear models were created correlating keyboard metadata collected from the week prior to the administration of the HDRS (17-item) and YMRS mood rating scores. Missing data were handled with pairwise deletion. Features extracted from the metadata were modeled as fixed effects. Observations were grouped by subject, with each subject having his or her own random intercept for his or her mood ratings.

Figure 1. Screenshot of the BiAffect keyboard (keyboard design derived from AnySoftKeyboard by Menny Evan-Danan and licensed under Apache License 2.0.).

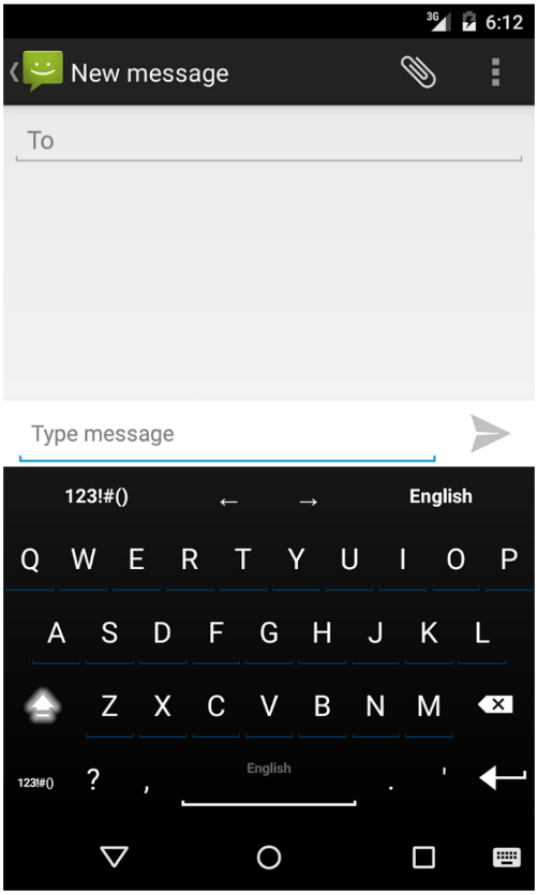


Table 1. Subject characteristics.

Characteristics	Value
Age in years, mean (SD)	48.67 (9.63)
Female gender, n (%)	8 (89)
Years of education, mean (SD)	14.00 (2.12)
Diagnosis, n (%)	
Bipolar I	1 (11)
Bipolar I with rapid cycling	4 (44)
Bipolar II, recurrent depressive episodes, with rapid cycling	4 (44)
Number of keystrokes, mean (SD)	69,627 (57,477)
Number of accelerometer readings, mean (SD)	740,099 (47,165)
Weeks of activity, mean (SD)	7.70 (0.70)
Initial HDRS[a]-17 item, mean (SD)	11.90 (6.17)
Final HDRS-17 item, mean (SD)	11.11 (5.49)
Initial YMRS[b], mean (SD)	7.56 (5.00)
Final YMRS, mean (SD)	6.67 (4.03)

[a]HDRS: Hamilton Depression Rating Scale.

[b]YMRS: Young Mania Rating Scale.

Table 2. Predictor variable definitions.

Predictor variable	Definition
Average interkey delay	The average time between keystrokes measured in seconds
Backspace ratio	Number of backspace keypresses divided by total keypresses
Autocorrect rate	Number of autocorrect events divided by total keypresses
Circadian baseline similarity	The cosine-based similarity between the hourly distribution of keypresses/week and the hourly distribution for the study period
Average accelerometer displacement	Square root of sum of squares of accelerometer displacement along each coordinate (x, y, z) averaged over the week (average of $\sqrt{x^2+y^2+z^2}$)
Average session length	Length of sessions in seconds averaged over the week
Session count	Number of sessions: A session begins when a keypress is initiated and ≥5 s has elapsed since the last key was pressed. A session ends when ≥5 s has elapsed since the last key was pressed.

Overall significance was assessed by using likelihood ratio tests comparing the null models that consisted of just the subject-level effect with full models consisting of the subject-level effect and metadata features. Because the mixed-effects model for the YMRS scores showed that the random effect was accounting for none of the variance of the YMRS scores, a fixed-effects ordinary least squares model was created instead (mixed-effects model log likelihood −64.621, Akaike Information Criterion 149.24, Bayesian Information Criterion 170.83; fixed-effects ordinary least squares model: log likelihood −64.621, Akaike Information Criterion 147.24, Bayesian Information Criterion 166.67). For the HDRS model, conditional and marginal R^2 values were calculated using the method specified by Nakagawa and Schielzeth [12], as implemented in the R package piecewiseSEM [13]. Using this method, the conditional R^2 is equal to the proportion of variance explained by both the fixed and random effects, and the marginal R^2 is equal to the proportion of the variance explained by the fixed effects alone. The *P* values of the model coefficients were calculated using Wald chi-square tests, as implemented in the R package car [14] for the HDRS model. For the YMRS model, overall significance was tested using an F-test and individual coefficient significance was determined with t-tests.

The fixed-effect variables included the average interkey typing delay, the average accelerometer displacement, the backspace and autocorrect rates (ie, the total number of each divided by the total number of keystrokes), the average length of each typing session in seconds, the total number of typing sessions, and the cosine similarity between each week's keypress activity and the total keypress activity of the study period (described further below). All aggregate variables were calculated for the week preceding each mood assessment. A session was defined as beginning with a keypress that occurs after 5 or more seconds have elapsed since the last keypress and ending when 5 or more seconds have elapsed between keypresses.

Models were created using the software package lme4 [15] for the R software environment version 3.3.3 [16].

Predictor Variables

The predictor variables were chosen based on the hypothesis that they map to key cognitive and behavioral domains affected by mania and depression. Table 2 provides definitions of each variable, and each domain and their corresponding variables are discussed in turn below.

Psychomotor Activity

As per the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), changes in psychomotor functioning are criteria for both major depressive and manic episodes [2]. Psychomotor activity is also a component of the clinician's ratings within HDRS and YMRS. We hypothesized that psychomotor activity (agitation and retardation) manifests in the accelerometer displacement and the average interkey delay. We predicted that increasing levels of psychomotor agitation lead to subjects holding their phones less stably, thus resulting in higher accelerometer displacement values. In the case of average interkey delay, it can be argued that increased levels of psychomotor agitation could lead to either a lower or higher delay. In the case of the former, higher levels of agitation would lead to a general speeding up of behaviors, including typing; however, it is also possible that while more agitation may lead to an increase in the amount of activity, the ability to effectively type will be impaired, leading to a higher interkey delay and possibly more use of backspace and autocorrect. In contrast, psychomotor retardation was hypothesized to manifest as a higher average delay.

Social Activity

The BiAffect app did not capture the context of keyboard activity; however, we hypothesized that increases in keyboard activity are likely associated with increased social activity consisting of both texting and social media usage and that more activity would be associated with higher YMRS scores and lower HDRS scores. There are mixed data on the role of social media use and depression, with some studies showing decreases [17] and others reporting increases in social media usage in both high school [18] and college [19] students.

Cognition

Impairments in attention and concentration are seen in both depressive and manic episodes, as described in the DSM-5 and previous studies [2]. Impulsivity and deficits in error correction have also been identified as features seen in manic episodes [20]. Variables that characterize concentration and cognition were hypothesized to include the average interkey delay, the backspace rate, and the autocorrect rate. It was hypothesized

that increased backspace rates indicated increased error correction and increased autocorrect rates indicated decreased error detection. Impaired concentration was hypothesized to manifest as increased interkey delay.

Diurnal Activity Patterns

Changes in sleep patterns are characteristic of both depressive and manic episodes. In the case of depression, this may take the form of insomnia or hypersomnia, whereas in the case of mania, there is typically a decreased need for sleep [2]. We expected that such changes in sleeping patterns would manifest as changes in phone typing activity. To characterize such changes, we created a cosine-based similarity feature of keypress activity. Cosine-based similarity is a frequently used technique in the field of machine learning and predictive algorithms to characterize the similarity between entities [21,22]. In our implementation, the distribution of keypress activity for a given week was defined as vector of 24 dimensions, with each dimension corresponding to an hour of the day. The value of the vector in each dimension was set equal to the number of keypresses in that hour. We then calculated the cosine of the angle between each week's vector and the vector representing activity for the entire study period. In this way, the more dissimilar a given week's pattern of activity was compared to the total activity, the lower the value of the cosine would be. It was hypothesized that more dissimilar weeks would correspond to higher HDRS and YMRS scores.

Results

Predictor Variable Summary Statistics

Summary statistics for each predictor variable are presented in Table 3.

Total Key Press Activity and Subject Demographics

No statistically significant correlations were found between total key counts and subjects' age (S=139.16, P=.68) and education levels (S=144.41, P=.60).

Prediction of Depression Symptoms

Likelihood ratio testing comparing the null model that consisted of just the subject-level random effect to the full model showed that the full model had superior fit (χ^2_7=17.6, P=.01; see Tables 4 and 5). The marginal R^2 (ie, the proportion of the variance explained by the metadata features) was 0.41, and the conditional R^2 (ie, the proportion of the variance explained by both the subject-level effect and the metadata features) was .63. Accelerometer displacement (P=.002), average interkey delay (P=.02), session count (P=.003), and the autocorrect rate (P=.004) were found to be positively correlated with the HDRS scores.

Prediction of Hypomania or Mania Symptoms

A multiple linear regression model was created that accounted for 34% of the variance of the natural logarithm of YMRS scores (multiple R^2=.34, $F_{7,56}$=4.08, root mean square error=.66, P=.001; Table 5). Accelerometer displacement (P=.003) was found to be positively correlated with YMRS scores, and the backspace rate (P=.01) was found to be negatively correlated.

Table 3. Variable summary statistics.

Statistics	Mean (SD)	Min	Max	Average number of observations per subject per week
Average accelerometer displacement (m/s^2)	9.56 (0.22)	9.07	9.99	104,076
Average interkey delay (s)	0.69 (0.36)	0.30	1.95	9780
Backspace rate	0.093 (0.050)	0.0070	0.27	982[a]
Autocorrect rate	0.10 (0.033)	0.011	0.14	1021[a]
Average session length (s)	19.63 (6.12)	6.90	31.51	—
Session count	241.13 (159.33)	9	804	—
Circadian baseline similarity	0.77 (0.17)	0.14	0.96	—
HDRS[b] 17-item	11.83 (6.29)	0	25	—
YMRS[c]	5.64 (4.87)	0	20	—
Natural log of YMRS	1.60 (0.82)	0	3.045	—

[a]Average number of backspace or autocorrect events.

[b]HDRS: Hamilton Depression Rating Scale.

[c]YMRS: Young Mania Rating Scale.

Table 4. Fixed effects estimates of regression models.

Scaled predictors	17-item Hamilton Depression Rating Scale		Natural log of Young Mania Rating Scale	
	Linear mixed-effects (95% CI)	*P* value	Ordinary least squares (95% CI)	*P* value
Average accelerometer displacement	3.20 (1.20 to 5.21)	.0017	0.39 (0.15 to 0.64)	.003
Average interkey delay	2.88 (0.42 to 5.35)	.022	0.13 (−0.19 to 0.44)	.44
Backspace ratio	−0.01 (−1.53 to 1.52)	.99	−0.30 (−0.53 to −0.070)	.014
Autocorrect rate	2.67 (0.87 to 4.47)	.0036	0.06 (−0.17 to 0.29)	.63
Average session length	−1.16 (−2.71 to 0.39)	.14	−0.04 (−0.24 to 0.16)	.68
Session count	2.18 (0.77 to 3.56)	.0025	−0.04 (−0.28 to 0.19)	.73
Circadian baseline similarity	0.34 (−1.07 to 1.75)	.64	0.03 (−0.22 to 0.27)	.83
Constant	11.77 (9.80 to 13.74)	<.001	1.60 (1.43 to 1.78)	<.001

Table 5. Summary of regression results.

Scaled predictors	17-item Hamilton Depression Rating Scale	Natural log of Young Mania Rating Scale
	Linear mixed-effects	Ordinary least squares
Observations	64	64
Multiple R^2	—	.34
Adjusted R^2	—	.26
Conditional R^2	.63	—
Marginal R^2	.41	—
Log likelihood	−179.65	—
Residual standard error	—	.71[a]
Chi-square statistic or F statistic	17.6[b]	4.1[c]

[a]df=56

[b]χ^2_7, P=.014.

[c]$F_{7,56}$, P=.0011.

Discussion

Principal Findings

Using only passively collected metadata, keystroke activity predicted both depressive and manic symptoms. The model to predict depression scores demonstrated greater explanatory capacity as shown by the larger proportion of variance explained by the model and the larger number of significant predictors.

Psychomotor Activity

Increased accelerometer activity was found to be positively correlated with both depression and mania scores. One possible explanation for the positive correlation with both scores is that the subjects in our study had more mildly agitated or irritable forms of depression or depression with mixed features rather than forms exhibiting psychomotor retardation.

Social Activity

In contrast to our hypothesis that decreased sessions would be predictive of higher depression, the overall number of sessions was actually positively correlated with depression. This may be a reflection of the dynamic between loneliness and withdrawal. Sessions from a phone can be seen as lower risk and can also include passive use of social media, such as viewing but not posting, enabling a feeling of connection and withdrawal. At least one study has demonstrated an association between increased usage of the internet more generally and depressive symptoms [23]. It is also worth noting that while the session count was positively correlated, the average session length was negatively correlated (although this predictor did not reach statistical significance, P=.15), suggesting that patterns of activity may be more relevant than the overall volume of activity.

Cognition

Impairments in executive function have been demonstrated more in individuals with bipolar disorder in depressed, manic, and euthymic states than in healthy controls [20], although it has also been shown that executive functioning may be especially impacted during manic states [24,25]. Interestingly, our depression and mania symptom models diverge in their relationships with respect to what we theorized would be the key features related to cognition: backspace and autocorrect rates. The increase in autocorrect rate with depression symptoms seems relatively straightforward. Here, the ability to concentrate

becomes impaired in more depressed states, and therefore, the rate of typing errors increases. What is less clear is why the backspace rate would be negatively correlated with mania symptoms without a concomitant positive correlation with the autocorrect rate. One possibility is that the lower backspace usage seen with higher mania scores reflects a phenomenon of less self-monitoring or impaired response inhibition with errors. Those with elevated mania do not trigger the autocorrect mechanism because their inputs are generally correctly spelled but often grammatically or semantically inappropriate words, fitting the profile of someone who keeps deleting what they type because it was impulsively entered.

Diurnal Activity Patterns

Because sleep disturbance is such a prominent aspect of mood disturbance, we were surprised that measurements that aimed to reflect diurnal variations in activity were not predictive of depressive or mania symptoms. With the assumption that the distribution for the entire observation period would approximate the subject's baseline, we expected that lower values of similarity would be correlated with higher depression and mania scores. The cosine similarity values did not reach statistical significance in both models. One possible explanation for this is that the period of observation was not long enough to establish actual baselines in the sense of encompassing activity through a variety of mood states, including euthymia, and that the distribution for the entire observation period for many subjects corresponded to a single mood state. Another important consideration is that while diurnal patterns of phone activity may be related to sleep, they are not identical.

Limitations and Future Directions

The limitations of this study include its sample size (relative to the model's complexity), sample characteristics that are probably not representative of a general population (ie, mostly women who have a high frequency of episodes), and the constraint of having subjects using study-issued phones. A larger study in which participants use their own phones is warranted in order to determine the generalizability of these findings. More data may also enable the creation of more sophisticated models with higher rates of prediction accuracy and reliability.

Unfortunately, there were fewer predictors of mania scores, and overall, this prediction was less accurate. Prediction of acute changes in mania may have stronger clinical implications, given the reduced tendency to seek treatment in mania generally. We suspect that primary reasons for the decreased prediction of mania are that our sample contained generally low mania scores and that both mania and hypomania elevations are often short and sporadically observed relative to longer and more stable episodes of depression. Rather than demonstrating correlates of mania per se, the mania model presented here might represent correlates of mixed or agitated depression.

Comparison with Prior Work

Prior studies have investigated the potential utility of various aspects of mobile phone activity as a means to diagnose mood states. Early studies focused on demonstrating the practicality of collecting self-reports of mood using mobile phones from patients [26,27]. While this approach may increase the facility with which such data are collected, it is still subject to the biases associated with self-reported data, potentially leading to spurious results [28]. More recent studies have focused on the validation of passive data collection methods and yielded encouraging results. Passive data features that have been demonstrated to correlate with mood ratings include physical movement [29,30], amount of phone usage [30], and frequency of calls and text messages with personal contacts [31].

The use of keystroke dynamics as a means to detect the emotion or mood of users is an active area of research in the field of affective computing, with most studies to date investigating the use of desktop keyboards [32]; however, there have been at least two studies that have examined the use of mobile phone keyboard dynamics as means to recognize user emotion. The first study was a 2-week pilot study based on the activity of a single user on Twitter, wherein the user was instructed to write a Tweet whenever he or she experienced certain emotions and to record the emotion from a preset selection of options. Using a Bayesian Network classifier, the investigators were able to achieve an overall classification accuracy of 67.52%, with the most important feature being typing speed [33]. The second study consisted of a larger sample of 22 subjects and was conducted over 3 weeks. It also presented users with a preset selection of options for emotions; although, in contrast to the first study, keyboard activity was recorded over all applications and the users were prompted to input their emotional state on a regular basis. Using a random forest model, the investigators were able to achieve an average classification accuracy of 84%, with the most important typing dynamic feature being typing speed [34].

Although the aforementioned studies measuring mobile phone keystroke dynamics sought to predict emotion rather than mood, we find the relative importance of typing speed as an important feature across their studies as well as our own to be of note. To the best of our knowledge, our study is the first effort to use passively collected mobile phone keyboard metadata features to predict mood disturbances in a clinical sample using clinically relevant measures.

Conclusions

Passively collected mobile phone keystroke dynamics may be a useful and important method to identify incipient mood processes in persons with bipolar disorder. The facility with which such data may be used to infer the presence and severity of mood disturbances may enable clinical providers to intervene earlier in their patients' mood episodes, as well as increase the number of patients a single provider can effectively manage. Models such as those presented here may also lead to a deeper understanding of these disorders by revealing novel behavioral traits associated with them.

Acknowledgments

The authors thank the staff and participants of the Heinz C Prechter Research Program. This work was funded by the Mood Challenge, a New Venture Fund program funded by the Robert Wood Johnson Foundation.

Authors' Contributions

AL had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

MM is listed as an inventor on US Patent US9685174 (mood monitoring of bipolar disorder using speech analysis).

References

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
2. American Psychiatric Association. Internet. Diagnostic and statistical manual of mental disorders (5th edition) URL: https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596 [accessed 2018-07-03] [WebCite Cache ID 70ec7MqC8]
3. Sajatovic M. Bipolar disorder: disease burden. Am J Manag Care 2005 Jun;11(3 Suppl):S80-S84 [FREE Full text] [Medline: 16097718]
4. Pew Research Center. US Smartphone Use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [WebCite Cache ID 70efwIJRs]
5. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
6. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May 12;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
7. McInnis MG, Assari S, Kamali M, Ryan K, Langenecker SA, Saunders EFH, Prechter Bipolar Clinical Research Collaborative. Cohort profile: the Heinz C. Prechter longitudinal study of bipolar disorder. Int J Epidemiol 2018 Feb 01;47(1):28-28n [FREE Full text] [doi: 10.1093/ije/dyx229] [Medline: 29211851]
8. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fourth edition, text revision. 2000. URL: https://dsm.psychiatryonline.org/doi/pdf/10.1176/appi.books.9780890420249.dsm-iv-tr [accessed 2018-07-08] [WebCite Cache ID 70lbHq4l5]
9. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23:56-62 [FREE Full text] [Medline: 14399272]
10. Young R, Biggs J, Ziegler V, Meyer D. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov;133:429-435. [Medline: 728692]
11. Doane D, Seward L. Measuring Skewness: A Forgotten Statistic? Journal of Statistics Education 2017 Aug 29;19(2):A [FREE Full text] [doi: 10.1080/10691898.2011.11889611]
12. Nakagawa S, Schielzeth H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods Ecol Evol 2012 Dec 03;4(2):133-142 [FREE Full text] [doi: 10.1111/j.2041-210x.2012.00261.x]
13. Lefcheck J. piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. Methods Ecol Evol 2016;7(5):579. [doi: 10.1111/2041-210X.12512]
14. Fox J, Weisberg S. An R Companion to Applied Regression Internet. Second Edition. Thousand Oaks, CA: Sage; 2011.
15. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Soft 2015;67(1):1 [FREE Full text] [doi: 10.18637/jss.v067.i01]
16. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
17. De Chourdry M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. 2013 Presented at: International Conference on Weblogs and Social Media; July 8-11, 2013; Cambridge, MA URL: https://www.microsoft.com/en-us/research/publication/predicting-depression-via-social-media/
18. Pantic I, Damjanovic A, Todorovic J, Topalovic D, Bojovic-Jovic D, Ristic S, et al. Association between online social networking and depression in high school students: behavioral physiology viewpoint. Psychiatr Danub 2012 Mar;24(1):90-93 [FREE Full text] [Medline: 22447092]
19. Steers M, Wickham R, Acitelli L. Seeing Everyone Else's Highlight Reels: How Facebook Usage is Linked to Depressive Symptoms. Journal of Social and Clinical Psychology 2014 Oct;33(8):701-731 [FREE Full text] [doi: 10.1521/jscp.2014.33.8.701]
20. Martínez-Arán A, Vieta E, Reinares M, Colom F, Torrent C, Sánchez-Moreno J, et al. Cognitive function across manic or hypomanic, depressed, and euthymic states in bipolar disorder. Am J Psychiatry 2004 Feb;161(2):262-270. [doi: 10.1176/appi.ajp.161.2.262] [Medline: 14754775]

21. Huang A. Similarity Measures for Text Document Clustering. 2008 Presented at: New Zealand Computer Science Research Student Conference 2008; 2008; New Zealand URL: http://nzcsrsc08.canterbury.ac.nz/site/proceedings/Individual_Papers/pg049_Similarity_Measures_for_Text_Document_Clustering.pdf
22. Singhal A. Modern information retrieval: A brief overview. IEEE Data Eng Bull 2001;24(4):43 [FREE Full text]
23. Katikalapudi R, Chellappan S, Montgomery F, Wunsch D, Lutzen K. Associating Internet Usage with Depressive Behavior Among College Students. IEEE Technol. Soc. Mag 2012 Dec;31(4):73-80 [FREE Full text] [doi: 10.1109/MTS.2012.2225462]
24. Dixon T, Kravariti E, Frith C, Murray RM, McGuire PK. Effect of symptoms on executive function in bipolar illness. Psychol Med 2004 Jul;34(5):811-821. [Medline: 15500302]
25. Ryan KA, Vederman AC, McFadden EM, Weldon AL, Kamali M, Langenecker SA, et al. Differential executive functioning performance by phase of bipolar disorder. Bipolar Disord 2012 Aug;14(5):527-536 [FREE Full text] [doi: 10.1111/j.1399-5618.2012.01032.x] [Medline: 22834461]
26. Courvoisier DS, Eid M, Lischetzke T, Schreiber WH. Psychometric properties of a computerized mobile phone method for assessing mood in daily life. Emotion 2010 Feb;10(1):115-124. [doi: 10.1037/a0017813] [Medline: 20141308]
27. Kuntsche E, Labhart F. Using Personal Cell Phones for Ecological Momentary Assessment. European Psychologist 2013 Jan;18(1):3-11 [FREE Full text] [doi: 10.1027/1016-9040/a000127]
28. Bauhoff S. Systematic self-report bias in health data: impact on estimating cross-sectional and treatment effects. Health Serv Outcomes Res Method 2011 Feb 9;11(1-2):44-53. [doi: 10.1007/s10742-011-0069-3]
29. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. Psychiatry Res 2014 Jun 30;217(1-2):124-127. [doi: 10.1016/j.psychres.2014.03.009] [Medline: 24679993]
30. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
31. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile Phone-Based Unobtrusive Ecological Momentary Assessment of Day-to-Day Mood: An Explorative Study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
32. Kolakowska A. A review of emotion recognition methods based on keystroke dynamics and mouse movements. 2013 Presented at: 6th International Conference on Human System Interactions (HSI); 2013; Sopot, Poland. [doi: 10.1109/HSI.2013.6577879]
33. Lee H, Choi YS, Lee S, Park IP. Towards Unobtrusive Emotion Recognition for Affective Social Communication. 2012 Presented at: 2012 IEEE Consumer Communications and Networking Conference (CCNC); 2012; Las Vegas, NV, USA.
34. Ghosh S, Ganguly N, Mitra B, De P. TapSense: Combining Self-Report Patterns and Typing Characteristics for Smartphone based Emotion Detection. 2017 Presented at: International Conference on Human-Computer Interaction with Mobile Devices and Services; 2017; Vienna, Austria URL: http://dl.acm.org/citation.cfm?doid=3098279.3098564

Abbreviations

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
HDRS: Hamilton Depression Rating Scale
YMRS: Young Mania Rating Scale
SMS: short messaging service

Edited by G Eysenbach; submitted 04.01.18; peer-reviewed by D Hidalgo-Mazzei, I Hungerbuehler, DA Rohani, C Karr; comments to author 24.03.18; revised version received 13.05.18; accepted 29.05.18; published 20.07.18.

Please cite as:
Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, McInnis M, Ajilore O, Nelson PC, Ryan K, Leow A
Predicting Mood Disturbance Severity with Mobile Phone Keystroke Metadata: A BiAffect Digital Phenotyping Study
J Med Internet Res 2018;20(7):e241
URL: http://www.jmir.org/2018/7/e241/
doi:10.2196/jmir.9775
PMID:30030209

Original Paper

Public Perception Analysis of Tweets During the 2015 Measles Outbreak: Comparative Study Using Convolutional Neural Network Models

Jingcheng Du[1*], BS; Lu Tang[2*], PhD; Yang Xiang[1], PhD; Degui Zhi[1], PhD; Jun Xu[1], PhD; Hsing-Yi Song[1], MPH; Cui Tao[1], PhD

[1]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States

[2]Department of Communication, College of Liberal Arts, Texas A&M University, College Station, TX, United States

[*]these authors contributed equally

Corresponding Author:
Cui Tao, PhD
School of Biomedical Informatics
The University of Texas Health Science Center at Houston
7000 Fannin Street
Houston, TX,
United States
Phone: 1 713 500 3981
Email: cui.tao@uth.tmc.edu

Abstract

Background: Timely understanding of public perceptions allows public health agencies to provide up-to-date responses to health crises such as infectious diseases outbreaks. Social media such as Twitter provide an unprecedented way for the prompt assessment of the large-scale public response.

Objective: The aims of this study were to develop a scheme for a comprehensive public perception analysis of a measles outbreak based on Twitter data and demonstrate the superiority of the convolutional neural network (CNN) models (compared with conventional machine learning methods) on measles outbreak-related tweets classification tasks with a relatively small and highly unbalanced gold standard training set.

Methods: We first designed a comprehensive scheme for the analysis of public perception of measles based on tweets, including 3 dimensions: discussion themes, emotions expressed, and attitude toward vaccination. All 1,154,156 tweets containing the word "measles" posted between December 1, 2014, and April 30, 2015, were purchased and downloaded from DiscoverText.com. Two expert annotators curated a gold standard of 1151 tweets (approximately 0.1% of all tweets) based on the 3-dimensional scheme. Next, a tweet classification system based on the CNN framework was developed. We compared the performance of the CNN models to those of 4 conventional machine learning models and another neural network model. We also compared the impact of different word embeddings configurations for the CNN models: (1) Stanford GloVe embedding trained on billions of tweets in the general domain, (2) measles-specific embedding trained on our 1 million measles related tweets, and (3) a combination of the 2 embeddings.

Results: Cohen kappa intercoder reliability values for the annotation were: 0.78, 0.72, and 0.80 on the 3 dimensions, respectively. Class distributions within the gold standard were highly unbalanced for all dimensions. The CNN models performed better on all classification tasks than k-nearest neighbors, naïve Bayes, support vector machines, or random forest. Detailed comparison between support vector machines and the CNN models showed that the major contributor to the overall superiority of the CNN models is the improvement on recall, especially for classes with low occurrence. The CNN model with the 2 embedding combination led to better performance on discussion themes and emotions expressed (microaveraging F1 scores of 0.7811 and 0.8592, respectively), while the CNN model with Stanford embedding achieved best performance on attitude toward vaccination (microaveraging F1 score of 0.8642).

Conclusions: The proposed scheme can successfully classify the public's opinions and emotions in multiple dimensions, which would facilitate the timely understanding of public perceptions during the outbreak of an infectious disease. Compared with conventional machine learning methods, our CNN models showed superiority on measles-related tweet classification tasks with a relatively small and highly unbalanced gold standard. With the success of these tasks, our proposed scheme and CNN-based

tweets classification system is expected to be useful for the analysis of tweets about other infectious diseases such as influenza and Ebola.

(J Med Internet Res 2018;20(7):e236) doi:10.2196/jmir.9413

KEYWORDS

convolutional neural networks; social media; measles; public perception

Introduction

Nearly 40 million cases of measles, caused by a highly contagious virus, lead to over 300,000 deaths worldwide every year [1]. In the United States, measles was officially declared to be eliminated in 2000 thanks to the successful nationwide administration of a 2-dose vaccination program [2]. However, recent years have seen the reemergence of measles outbreaks in the United States. The most recent large-scale measles outbreak occurred in early 2015 with a high concentration of cases in California [3]. Researchers believe that increasing rates of vaccination refusal and undervaccination have made the public more vulnerable to this potentially deadly disease [4].

During an outbreak of an infectious disease such as measles, responsible public health agencies need to send out timely messages to the public during different stages of the crisis [5]. For instance, the Centers for Disease Control and Prevention (CDC) has adopted a 5-stage model of crisis and emergency risk communication, including precrisis, initial event, maintenance, resolution, and evaluation [5]. Prompt understanding of the public's perceptions will allow public health agencies to respond to people's attitudes, emotions, and needs in real time instead of relying on a predetermined timeline based on stages. Using traditional methods such as surveys to study public perceptions during an infectious disease outbreak is both costly and time-consuming [4,6].

Social media have been increasingly used by the general public, patients, and health professionals to communicate about health-related issues [7]. Researchers have studied social media content for drug adverse events detection [8,9], assessment of public opinion about health-related issues such as vaccination [10-13], and infectious disease outbreak surveillance [6,14,15]. Twitter, one of the largest public social media in the world, provides unique insights into how the public responds to an infectious disease outbreak as users, in real time, share information about the outbreak, talk about their personal experiences, argue over the necessity and safety of vaccination, and express a wide range of emotions. Examining Twitter content can provide an immediate assessment of the public's response and will allow public health professionals to adapt their messages to communicate with the public more effectively.

Many studies have used Twitter to assess various public health topics. However, most of the studies thus far have focused on analyzing the frequency of postings rather than on understanding post contents [16]. There is an increasing need to develop automatic and scalable approaches for the accurate understanding of the high volume of Twitter posts. Recent advances in machine learning and natural language processing (NLP) technologies allow for the stringent analysis of large amounts of Twitter posts. However, compared to texts in other domains, Twitter text has very distinctive characteristics such as very short text, unique Twitter language and structures, etc. For some health-related topics, there also exists the unbalanced class distribution issue (certain classes are much more frequent than other classes), which can further erode the performance of NLP models [10,13]. To improve performance on health-related Twitter datasets, substantial time and effort on feature engineering [10,17,18] is needed for conventional machine-learning algorithms, including support vector machines (SVMs), k-nearest neighbors (KNNs), etc.

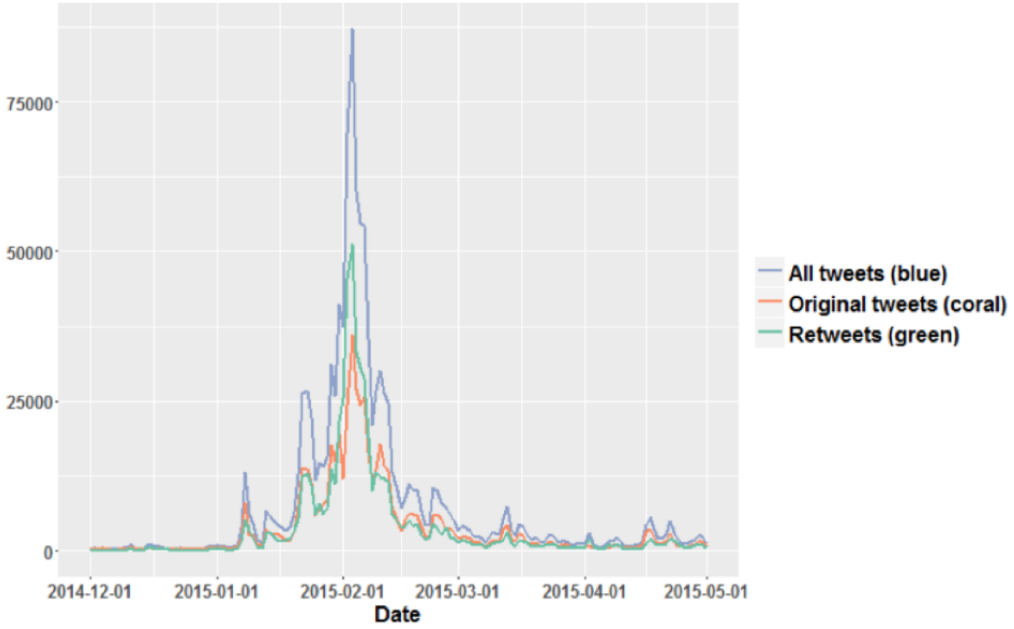
Compared to conventional machine learning algorithms, neural network models are advantageous because they have saved significant time on task-specific features engineering, achieved higher performance, and are scalable to large applications [19]. Some recent works applied neural network models to social media to understand public perceptions and behaviors. For instance, Lima et al [20] investigated the use of a multilayer perceptron neural network to classify personality from Twitter. Huynh et al [21] and Coco et al [22] proposed a deep neural network model to identify adverse drug reactions from Twitter data. Kendra [23] used a 5-layer neural network to characterize the discussion about antibiotics on Twitter. Bian et al [24] applied a convolutional neural network model to perform sentiment analysis on layperson's tweets. Zhao et al [25] proposed a semisupervised deep learning for influenza epidemic simulation. However, to our best knowledge, little work has been done to study public perceptions of infectious diseases and vaccinations on Twitter using neural network models.

Methods

Data Collection

All tweets including the word "measles" posted between December 1, 2014, and April 30, 2015, were purchased and downloaded from DiscoverText.com. This time frame was chosen because the unidentified Patient Zero of this outbreak visited the Disneyland theme park in California in December 2014. The first few suspected cases of measles were reported on January 5, 2015, and the last case was reported on March 2, 2015. CDC officially declared the outbreak to be over on April 17, 2015 [26]. A total of 1,154,156 tweets were collected. The number of tweets collected during the time frame can be seen in Figure 1.

Figure 1. Frequency of measles-related tweets by date and type.



Gold Standard Annotation

In order to understand measles-related contents on Twitter comprehensively, we created an annotation scheme containing 3 dimensions: *discussion themes*, *emotions expressed*, and *attitude toward vaccination*. The coding schemes *discussion themes* and *emotions expressed* were adapted based on Chew and Eysenbach [6], while the coding scheme *attitude toward vaccination* was created by the authors inductively. For *discussion themes*, 5 themes were identified: resources (news update about the outbreak, medical information about prevention, treatment, symptoms of measles), personal experience (direct or indirect experiences about measles), personal opinions and interests, questions, and other (unrelated to measles). *Emotions expressed* was categorized into 5 types: humor or sarcasm, positive emotion (relief and downplayed risk), anger, concern, and not applicable. The data collection was based on the keyword measles; however, debate about vaccines emerged in a large percentage of tweets collected. Hence, we took this opportunity to measure how public opinion changed over time during a measles outbreak. *Attitude toward vaccination* was categorized into 3 groups: pro (provaccination), against (antivaccination), and not applicable (no attitude). See Figure 2 for a visual representation of the 3 dimensions and categories within each dimension.

Two coders manually coded 0.1% of all tweets selected through systematic sampling. The first tweet was identified using a random number generator. After this, every 1000th tweet was selected in the sample. The Cohen kappa intercoder reliability values for the 3 dimensions were 0.78, 0.72, 0.80, respectively. Afterward, the 2 coders discussed their results to resolve discrepancies.

Neural Network Classification System

Data Cleaning

The vocabulary used on Twitter is very different from the general English vocabulary. User names, URLs, and hashtags need to be normalized. We first replaced tokens containing all capital letters with the lowercase of the token with string "<ALLCAPS>". Then all URLs were replaced with string "<URL>". Twitter user names (eg, @twitter) were then replaced with string "<USER>". All numbers were replaced with string "<NUMBER>". All hashtags were separated into tokens by uppercase letters (eg, we replace "#VaccineWork" with "<HASHTAG> Vaccine Work"). Afterwards, all tweets were converted to lowercase. Our tweets preprocessing process was based on the Stanford GloVe tweets preprocessing script [27]. An example illustrating the tweet preprocessing step is shown below:

Raw tweet text: "RT @KTLA: #BREAKING: At least 9 measles cases linked to visits to @Disneyland from Dec. 15-20 http://t.co/1GRlwFhPgv http://t.co/3Nl15jmqAE"

Cleaned tweet text: "rt <allcaps> <user>: breaking: at least <number> measles cases linked to visits to <user> from dec. <number> <number> <url> <url>"

Convolutional Neural Networks

Commonly used in various computer vision tasks [28], convolutional neural networks (CNNs) have demonstrated excellent performance in the NLP field, including different text classification tasks [29-32]. We extended the classic CNN framework for sentence classification proposed by Kim [29] by using combination generic Twitter embedding and target domain Twitter embedding [33]. Details of our CNN system architecture can be seen in Figure 3. We cleaned the tweets following the data cleaning step. Then each token of the tweets was mapped to 2 high-dimension representations through 2 word embeddings: generic tweets embedding and target domain tweets embedding. Both embeddings were fine-tuned during the training process.

Figure 2. Measles tweets annotation scheme for different dimensions.

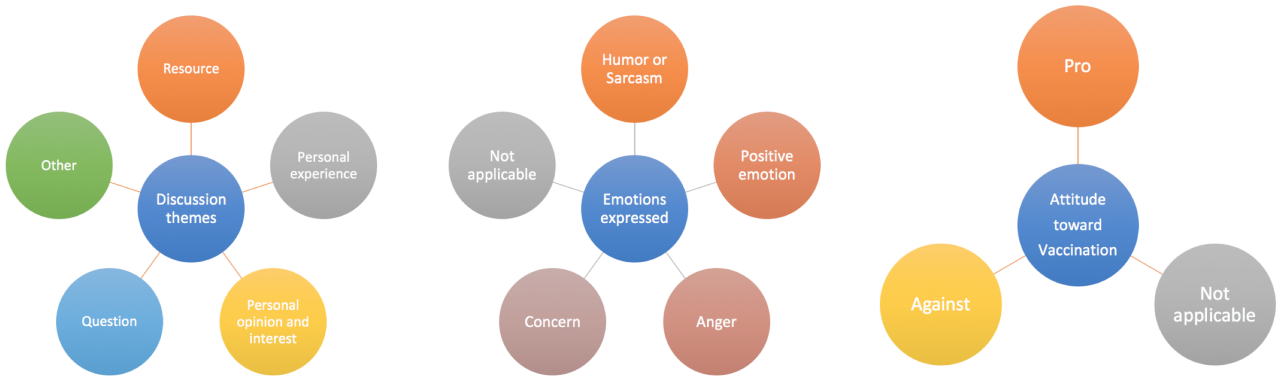
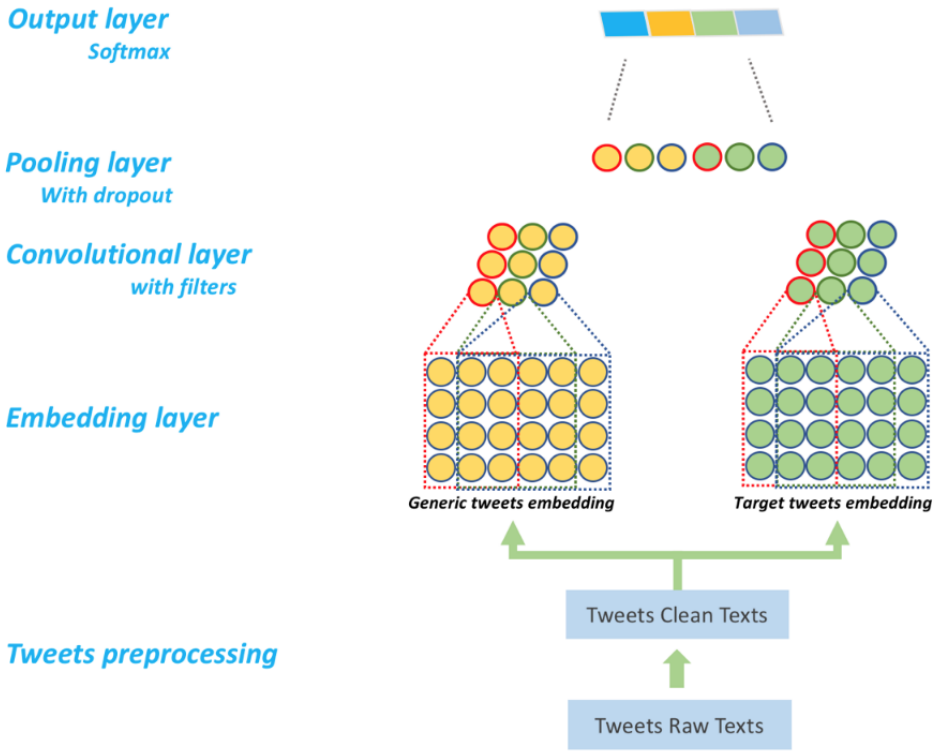


Figure 3. System architecture for measles-related tweets classification using convolutional neural networks.



We used 3 filters of size 3, 4, and 5 to generate the convolutional layer on each embedding. The feature maps generated by filters from each embedding were concatenated and fed to the pooling layer. We adopted max-pooling strategy with a dropout rate at 0.5 on the pooling layer. The output layer consisted of different classes for each dimension. This CNN system was built based on the Python and Tensorflow libraries [34].

Tweets Word Vector Embedding

For generic tweets embedding, we used pretrained GloVe tweets embedding from Stanford. GloVe is an unsupervised learning algorithm developed by Pennington et al [35] to obtain vector representations for words. GloVe tweets word vectors were trained on 2 billion tweets and 27 billion tokens [35] and have been widely used in different Twitter-related NLP tasks [31,36,37]. For target domain embedding, we trained a tweets embedding from our own measles-related tweets corpus (1,154,156 tweets) using the same GloVe algorithm. We tested different numbers of embedding dimensions in our preexperiments. The tweets word embedding in dimension 200 achieved the best performance for our tasks.

Experiments

For the CNN-based framework, we performed the following experiments: (1) use of pretrained GloVe tweets embedding only, (2) use of tweets measles embedding only, and (3) use of a combination of the pretrained GloVe tweets embedding and measles tweets embedding. For the use of 1 embedding only, we just used 1 channel of the proposed framework. We chose 4 popular machine learning models for comparison as our baselines: KNN [38], naïve Bayes [39], SVM [40], and random forest [41]. For SVM, a radial basis function kernel was used. We followed the same tweet cleaning steps and extracted n-grams as the feature for these traditional machine learning models. The Waikato Environment for Knowledge Analysis library was used to train and test these models [42]. We also evaluated the bidirectional long short-term memory (Bi-LSTM), which has achieved state-of-the-art performance in many

classification and sequence labeling tasks [43,44], for tweets classifications. The input of the Bi-LSTM is the pretrained GloVe tweets embedding (dimension: 200). We conducted these experiments on all 3 dimensions for public perceptions on measles.

System Evaluation

We leveraged a 10-fold cross-validation to evaluate the performances of these models for each classification task. Standard metrics including precision, recall, and F1 score were calculated for each class. We also calculated the microaveraging F score and macroaveraging F score to evaluate their performance on each classification task. For microaveraged score, we summed up all the individual true positives, false positives, and false negatives. For macroaveraged score, we took the average of the F1 score of different categories.

Ethical Approval

This study received institutional review board approval from the Committee for the Protection of Human Subjects at the University of Texas Health Science Center at Houston. The reference number is HSC-SBMI-16-0291.

Results

Gold Standard Description

In total, 1151 tweets were annotated. Class distributions were highly unbalanced for all 3 tasks (Table 1). In terms of *discussion themes*, nearly two-thirds (718/1151, 62.38%) of tweets were categorized as resources (ie, outbreak update or medical information about measles). Less than one-third (344/1151, 29.89%) of the tweets were about users' personal opinions and interests. Only 1.82% (21/1151) of the tweets discussed personal experience with measles, and 1.73% (20/1151) asked questions. For *emotions expressed*, 79.84% (919/1151) of tweets were categorized as expressing concern. Humor or sarcasm was found in 9.47% (109/1151) of the tweets. Positive emotion and anger were found in 3.38% (39/1151) and 3.04% (35/1151) of the tweets, respectively. Finally, in terms of *attitude toward vaccination*, the majority of the tweets (913/1151, 79.32%) did not express any opinion about vaccination, 17.55% (202/1151) of tweets were provaccination and 3.13% (36/1151) were antivaccination.

Overall Comparison of Convolutional Neural Network Models With Conventional Models

Comparison of the performances of CNN models and 4 machine learning models on the 3 dimensions can be seen in Table 2. As shown, CNN-based models have better performance than other conventional machine learning models or the Bi-LSTM model. The CNN model with the combination of 2 embeddings achieved the best performance on *emotions expressed* and the highest macroaveraging F score on *discussion themes*. The CNN model with Stanford embedding had the highest microaveraging F score on *discussion themes* and achieved the best performance on *attitude toward vaccination*. The CNN with measles embedding achieved relatively high microaveraging F score on *emotions expressed* and *attitude toward vaccination*. The Bi-LSTM model had the worst performance among neural network models, probably due to the limited size of training data.

Table 1. Class distribution in the gold standard for 3 dimensions.

Dimension and class	Tweets, n (%)
Discussion themes	
Resource	718 (62.4)
Personal experience	21 (1.8)
Personal opinions and interest	344 (29.9)
Question	20 (1.7)
Other	48 (4.2)
Emotions expressed	
Humor or sarcasm	109 (9.5)
Positive emotion	39 (3.4)
Anger	35 (3.0)
Concern	919 (79.8)
Not applicable	49 (4.3)
Attitude toward vaccination	
Pro	202 (17.6)
Against	36 (3.1)
Not applicable	913 (79.3)

Table 2. Ten-fold cross-validation results of neural network models and 4 conventional machine learning models on 3 dimensions. Italics indicate best performance in that class.

Model	Microaveraging F score			Macroaveraging F score		
	Discussion themes	Emotions expressed	Attitude toward vaccination	Discussion themes	Emotions expressed	Attitude toward vaccination
KNN[a]	0.5143	0.6977	0.8129	0.3223	0.4074	0.5114
Naïve Bayes	0.6811	0.7767	0.7171	0.4101	0.4814	0.5343
Random forest	0.7350	0.8393	0.8085	0.4243	0.4393	0.5356
SVM[b]	0.7696	0.8365	0.8211	0.3917	0.4269	0.5345
Bi-LSTM[c]	0.7315	0.8271	0.7958	0.2899	0.3730	0.4358
CNN_M[d]	0.7533	0.8480	0.8355	0.4282	0.4849	0.5871
CNN_S[e]	*0.7897*	0.8575	*0.8642*	0.4158	0.5419	*0.6629*
CNN_M+S[f]	0.7811	*0.8592*	0.8254	*0.4611*	*0.5591*	0.6078

[a]KNN: k-nearest neighbor.

[b]SVM: support vector machines.

[c]Bi-LSTM: bidirectional long short-term memory.

[d]CNN_M: convolutional neural network using the measles tweets embedding.

[e]CNN_S: convolutional neural network using the pretrained GloVe tweets embedding from Stanford.

[f]CNN_M+S: convolutional neural network using the combination of pretrained GloVe tweets embedding and measles tweets embedding.

As shown in Table 2, among the conventional machine learning models, SVM generally performed the best on all 3 dimensions. In order to further compare the performances of CNN models on each class and try to improve the overall performance, we then calculated and compared the precision, recall, and F score of SVM, the CNN model with Stanford GloVe tweets embedding only, and the CNN model with the combination of generic and target domain embedding.

Detailed Comparison of Convolutional Neural Network Models With Support Vector Machines on 3 Dimensions

Table 3 shows the comparison of SVM and CNN models on *discussion themes*. For precision score, the CNN with GloVe tweets embedding achieved better performance on classes with larger numbers of tweets (resources and personal opinions and interest). The CNN with the combination of 2 embeddings achieved better performance on classes with very limited numbers of tweets (ie, questions). For recall score, the CNN model with either Stanford embedding or the combination of 2 embeddings greatly improved the recall of the classes with relatively fewer tweets such as personal opinions and interests and questions, while SVM had slightly better performance on resources. The improvement of recall score greatly contributed to the improvement on the F score. Unfortunately, for the class personal experience, none of the models could identify any tweets correctly.

The comparison of SVM and the CNN models on *emotions expressed* can be seen in Table 4. CNN models achieved higher precision scores on classes with fewer cases, including anger and not applicable, while SVM performed better on humor or sarcasm. For recall and F1 score, CNN models with either Stanford embedding or the combination of 2 embeddings performed well on all classes. In general, the CNN with the combination of 2 embeddings had better performance for more categories than the CNN with Stanford embedding only.

For dimension 3, *attitude toward vaccination*, the overall comparison between the CNN models and SVM can be seen in Table 5. Both CNN models outperformed SVM in most of the categories, and the CNN model with Stanford embedding achieved better performance in most of the categories. Specifically, for precision score, SVM performed better on class pro, while the CNN models did better on class against and not applicable. The CNN with the combination of 2 embeddings achieved the highest precision score on against. In terms of recall, the CNN models performed much better on the classes with very small numbers of tweets (ie, pro and against), while SVM did better on the class not applicable. As for F1 score, the CNN with Stanford embedding performed the best, and SVM performed the worst on all 3 classes.

Table 3. Detailed precision, recall, and F score of each class for *discussion themes*. Italics indicate best performance in that class.

Class	Precision			Recall			F1 score		
	SVM[a]	CNN_M+S[b]	CNN_S[c]	SVM	CNN_M+S	CNN_S	SVM	CNN_M+S	CNN_S
Resource (n=718)	0.7907	0.8119	*0.8172*	*0.9471*	0.9318	0.9401	0.8619	0.8677	*0.8744*
Personal experience (n=21)	0	0	0	0	0	0	0	0	0
Personal opinions and interest (n=344)	0.7021	0.6984	*0.7231*	0.5773	0.6192	*0.6453*	0.6336	0.6564	*0.6820*
Question (n=20)	0	0.5	0	0	0.0500	0	0	*0.0909*	0
Other (n=48)	0.8750	0.8421	*0.8571*	0.1458	*0.3333*	0.2500	0.2500	*0.4776*	0.3871

[a]SVM: support vector machines.

[b]CNN_M+S: convolutional neural network using the combination of pretrained GloVe tweets embedding and measles tweets embedding.

[c]CNN_S: convolutional neural network using the pretrained GloVe tweets embedding from Stanford.

Table 4. Detailed precision, recall and F scores of each class for *emotions expressed*. Italics indicate best performance in that class.

Class	Precision			Recall			F1 score		
	SVM[a]	CNN_M+S[b]	CNN_S[c]	SVM	CNN_M+S	CNN_S	SVM	CNN_ M+S	CNN_S
Humor or sarcasm (n=109)	*1*	0.9388	0.8909	0.3486	0.4220	*0.4495*	0.5170	0.5823	*0.5976*
Positive emotion (n=39)	*1*	*1*	*1*	0.0513	*0.1538*	0.1282	0.0967	*0.2667*	0.2273
Anger (n=35)	0	*1*	0.6667	0	0.0286	*0.0571*	0	0.0556	*0.1053*
Concern (n=919)	0.8312	0.8538	*0.8550*	0.9069	*0.9978*	0.9946	0.9069	*0.9202*	0.9195
Not applicable (n=49)	0.7500	*0.9048*	0.8947	0.2105	*0.3878*	0.3469	0.2105	*0.5429*	0.5000

[a]SVM: support vector machines.

[b]CNN_M+S: convolutional neural network using the combination of pretrained GloVe tweets embedding and measles tweets embedding.

[c]CNN_S: convolutional neural network using the pretrained GloVe tweets embedding from Stanford.

Table 5. Detailed precision, recall, and F score of each class for *attitude toward vaccination*. Italics indicate best performance in that class.

Class	Precision			Recall			F1 score		
	SVM[a]	CNN_M+S[b]	CNN_S[c]	SVM	CNN_M+S	CNN_S	SVM	CNN_M+S	CNN_S
Pro (n=202)	*0.7917*	0.6458	0.7554	0.1919	0.3069	*0.5198*	0.3089	0.4161	*0.6158*
Against (n=36)	0.6667	*1*	0.8571	0.0556	*0.1667*	*0.1667*	0.1026	*0.2857*	0.2791
Not applicable (n=913)	0.8228	0.8408	*0.8794*	*0.9890*	0.9660	0.9682	0.8982	0.8991	*0.9216*

[a]SVM: support vector machines.

[b]CNN_M+S: convolutional neural network using the combination of pretrained GloVe tweets embedding and measles tweets embedding.

[c]CNN_S: convolutional neural network using the pretrained GloVe tweets embedding from Stanford.

Discussion

Principal Contributions

This study makes 2 primary contributions. First, we designed and implemented a comprehensive scheme for the public perception analysis of measles-related tweets, including *discussion themes*, *emotions expressed*, and *attitude toward vaccination*. We manually curated a gold standard set that contains 1151 tweets annotated according the scheme. The tweets were sampled from all measles-related tweets during the most recent measles outbreak in the United States in 2015. Based on the annotation results, we believe the scheme can successfully classify the public's opinions and emotions. Second, we designed and implemented CNN models on the classification tasks of measles-related tweets and investigated their performance compared to traditional machine learning models through a comprehensive comparison on the small-scale tweets corpus with highly unbalanced class distribution.

Principal Findings

In classifying measles-related tweets in terms of *discussion themes*, *emotions expressed*, and *attitude toward vaccination*, different classifiers were better suited for different tasks. However, the CNN models achieved better overall performance on all 3 tasks compared to conventional machine learning

algorithms. A detailed comparison of the CNN models and SVM showed that the CNN models were able to improve performance on nearly all classes for all 3 dimensions. The major contributor to the overall performance boost is the improvement on recall, especially for the classes with fewer cases than average. The CNN model with the combinations of 2 embeddings led to better performance on *discussion themes* and *emotions expressed*, while the CNN model with Stanford embedding achieved best performance on *attitude toward vaccination*. A common obstacle of deep neural network-based models is the need for a large training dataset. However, for a disease-related tweets classification task like ours, the results show that CNN models can perform better than conventional machine learning models even on a training dataset with only 1151 labeled tweets.

Limitations and Future Directions

Although the CNN models can greatly increase the performance for most of the classes with few cases, for some minor classes with extremely low numbers of cases such as personal experience in *discussion themes*, the CNN models are just as powerless as conventional models. Further examination of the prediction results shows that many tweets in the minor classes were incorrectly classified into major classes. For example, the tweets in personal experience were either classified as resources or personal opinions and interest. For against in *attitude toward vaccination*, the majority of the tweets were classified as not applicable, which takes up to 79% of the labeled data. The highly unbalanced class distribution is a major challenge for both conventional machine learning methods and neural network methods. Since the current gold standard training set is relatively small, we plan to collect and annotate more related tweets (especially the tweets belonging to smaller classes) to build a larger labeled dataset. We believe performance could be improved by using a larger labeled training dataset.

Future research could take a few directions. Additional hyperparameter tuning (ie, activation functions selection, pooling strategies) can also improve the performance on the disease-related tweets classification tasks. In addition, although the Bi-LSTM model doesn't work well on our tasks (probably due to the limited training data size), other recurrent neural network-based frameworks such as attentive Bi-LSTM [45] may lead to better performance, especially as the size of the training data increases. The improved models can be used to automatically predict the labels of the measles tweets, which will facilitate the analysis of large scale public perceptions about measles as well as other infectious diseases. Some unsupervised machine learning methods can also be used to explore the major discussion topics from the measles-related tweets dataset, such as topic modeling methods [46,47], as it can save the effort of annotation.

Conclusion

Timely understanding of public perceptions during the outbreak of an infectious disease such as measles will allow public health agencies to adapt their messages to address the needs, concerns, and emotions of the public. In order to understand the contents of Twitter text regarding measles and vaccination, we designed a classification scheme that contains *discussion themes*, *emotions expressed*, and *attitude toward vaccination* for measles-related tweets. A gold standard containing 1151 tweets was collected and manually annotated according to the classification scheme. CNN models have been evaluated to classify tweets into different classes for different tasks. A comparative study was done to evaluate the performance of CNN models in comparison to 4 conventional machine learning models as well as a Bi-LSTM model. The CNN models had improved performance on classification of themes, emotions, and attitude from the highly unbalanced measles-related tweets dataset. The CNN models presented in the paper can be applied on large-scale tweets datasets. Our proposed scheme and CNN-based tweets classification system for the public perception analysis on Twitter toward measles disease can be used for other infectious diseases such as influenza and Ebola.

Acknowledgments

This research was partially supported by the National Library of Medicine of the National Institutes of Health under award number R01LM011829, the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI130460, and the UTHealth Innovation for Cancer Prevention Research Training Program Pre-Doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant #RP160015). This study was also partially supported by a University of Alabama System's Collaborative Grant.

Conflicts of Interest

None declared.

References

1. Griffin D, Oldstone M. Measles: History and Basic Biology. Berlin: Springer Science & Business Media; 2008.
2. Katz S, Hinman A. Summary and conclusions: measles elimination meeting, 16-17 March 2000. J Infect Dis 2004 May 01;189 Suppl 1:S43-S47. [doi: 10.1086/377696] [Medline: 15106088]
3. Zipprich J, Winter K, Hacker J, Xia D, Watt J, Harriman K. Measles outbreak—California, December 2014-February 2015. MMWR Morb Mortal Wkly Rep 2015;64(6):154 [FREE Full text]
4. Dredze M, Broniatowski D, Smith M, Hilyard K. Understanding vaccine refusal: why we need social media now. Am J Prev Med 2016 Apr;50(4):550-552 [FREE Full text] [doi: 10.1016/j.amepre.2015.10.002] [Medline: 26655067]

5. Reynolds BW. Crisis and emergency risk communication as an integrative model. J Health Commun 2005;10(1):43-55. [doi: 10.1080/10810730590904571] [Medline: 15764443]
6. 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]
7. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
8. Yates A, Goharian N. ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. 2013 Presented at: Eur Conf Inf Retr; 2013; Moscow p. 816-819.
9. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug Saf 2014 May;37(5):343-350 [FREE Full text] [doi: 10.1007/s40264-014-0155-x] [Medline: 24777653]
10. Du J, Xu J, Song H, Liu X, Tao C. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. J Biomed Semantics 2017 Mar 03;8(1):9 [FREE Full text] [doi: 10.1186/s13326-017-0120-6] [Medline: 28253919]
11. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong M, Dunn AG. Using social connection information to improve opinion mining: identifying negative sentiment about HPV vaccines on Twitter. Stud Health Technol Inform 2015;216:761-765. [Medline: 26262154]
12. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
13. Du J, Xu J, Song H, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):69 [FREE Full text] [doi: 10.1186/s12911-017-0469-6] [Medline: 28699569]
14. Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. 2010 Presented at: Proc First Work Soc Media Anal; 2010; Washington p. 115-122.
15. Schmidt C. Trending now: using social media to predict and track disease outbreaks. Environ Health Perspect 2012 Jan;120(1):A30-A33 [FREE Full text] [doi: 10.1289/ehp.120-a30] [Medline: 22214548]
16. Myneni S, Fujimoto K, Cobb N, Cohen T. Content-driven analysis of an online community for smoking cessation: integration of qualitative techniques, automated text analysis, and affiliation networks. Am J Public Health 2015 Jun;105(6):1206-1212. [doi: 10.2105/AJPH.2014.302464] [Medline: 25880942]
17. Lopez M, Kalita J. Deep learning applied to NLP. arXiv Prepr arXiv170303091 2017:1 [FREE Full text]
18. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. arXiv Prepr arXiv170802709 2017:1 [FREE Full text]
19. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: MIT Press; 2016.
20. Lima ACES, de Castro LN. A multi-label, semi-supervised classification approach applied to personality prediction in social media. Neural Netw 2014 Oct;58:122-130. [doi: 10.1016/j.neunet.2014.05.020] [Medline: 24969690]
21. Huynh T, He Y, Willis A, Uger S. Adverse drug reaction classification with deep neural networks. Proc Coling Tech Papers 2016:877-887 [FREE Full text]
22. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. J Am Med Inform Assoc 2017 Jul 01;24(4):813-821. [doi: 10.1093/jamia/ocw180] [Medline: 28339747]
23. Kendra RL, Karki S, Eickholt JL, Gandy L. Characterizing the discussion of antibiotics in the Twittersphere: what is the bigger picture? J Med Internet Res 2015;17(6):e154 [FREE Full text] [doi: 10.2196/jmir.4220] [Medline: 26091775]
24. Bian J, Zhao Y, Salloum RG, Guo Y, Wang M, Prosperi M, et al. Using social media data to understand the impact of promotional information on laypeople's discussions: a case study of Lynch syndrome. J Med Internet Res 2017 Dec 13;19(12):e414 [FREE Full text] [doi: 10.2196/jmir.9266] [Medline: 29237586]
25. Zhao L, Chen J, Chen F, Wang W, Lu C, Tech V. SimNest: social media nested epidemic simulation via online semi-supervised deep learning. Proc IEEE Int Conf Data Min 2015 Nov:639-648. [doi: 10.1109/ICDM.2015.39.SimNest]
26. California Department of Public Health. 2015 Apr 17. California measles surveillance update URL: https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/Immunization/IMM-MeaslesUpdate2015-04-17.pdf [accessed 2018-04-02] [WebCite Cache ID 6yMpyCvVq]
27. Script for preprocessing tweets. URL: https://nlp.stanford.edu/projects/glove/preprocess-twitter.rb [accessed 2018-04-02] [WebCite Cache ID 6yMqCCzuH]
28. LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision. 2010 Presented at: Circuits Syst (ISCAS), Proc IEEE Int Symp; 2010; Paris p. 253-256.
29. Kim Y. Convolutional neural networks for sentence classification. arXiv Prepr arXiv14085882 2014:1 [FREE Full text]
30. Dos SC, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts. Proc Coling Tech Papers 2014:69-78 [FREE Full text]

31. Nakov P, Ritter A, Rosenthal S, Sebastiani F, Stoyanov V. SemEval-2016 task 4: Sentiment analysis in Twitter. 2016 Presented at: Proc SemEval; 2016; San Diego p. 1-18.
32. Conneau A, Schwenk H, Le Cun Y, Barrault L. Very deep convolutional networks for text classification. arXiv Prepr arXiv160601781 2016:1 [FREE Full text]
33. Limsopatham N, Collier N. Modelling the combination of generic and target domain embeddings in a convolutional neural network for sentence classification. Assoc Computational Linguistics 2016:136-140. [doi: 10.17863/CAM.4667]
34. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv Prepr arXiv160304467 2016:1 [FREE Full text]
35. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. 2014 Presented at: Proceedings of the conference on empirical methods in natural language processing; 2014; Doha p. 1532-1543.
36. Astudillo R, Amir S, Ling W, Martins B, Silva M, Trancoso I, et al. Inesc-id: a regression model for large scale twitter sentiment lexicon induction. 2015 Presented at: Proc 9th Int Work Semant Eval (SemEval); 2015; Lisbon p. 613-618.
37. Yamada I, Takeda H, Takefuji Y. Enhancing named entity recognition in twitter messages using entity linking. 2015 Presented at: Proceedings of the ACL 2015 Workshop on Noisy User-generated Text; 2015; Beijing p. 136-140.
38. Peterson L. K-nearest neighbor. Scholarpedia 2009;4(2):1883. [doi: 10.4249/scholarpedia.1883]
39. Murphy K. Naive Bayes Classifiers [dissertation]. Vancouver: University of British Columbia; 2006.
40. Hearst M, Dumais S, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst their Appl IEEE 1998;13(4):28. [doi: 10.1109/5254.708428]
41. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2(3):18-22 [FREE Full text]
42. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The WEKA data mining software: an update. ACM SIGKDD Explor Newsl 2009;11(1):10-18 [FREE Full text]
43. Dernoncourt F, Lee J, Uzuner O, Szolovits P. De-identification of patient notes with recurrent neural networks. J Am Med Informatics Assoc 2016:156. [doi: 10.1093/jamia/ocw156]
44. Zhang S, Zheng D, Hu X, Yang M. Bidirectional long short-term memory networks for relation classification. 2015 Presented at: Proc 29th Pacific Asia Conf Lang Inf Comput; 2015; Shanghai p. 73-78.
45. Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, et al. Attention-based bidirectional long short-term memory networks for relation classification. 2016 Presented at: Proc 54th Annu Meet Assoc Comput Linguist (Volume 2); 2016; Berlin p. 207-212. [doi: 10.18653/v1/P16-2034]
46. Yu Z, Johnson TR, Kavuluru R. Phrase based topic modeling for semantic information processing in biomedicine. Proc Int Conf Mach Learn Appl 2013 Dec:440-445 [FREE Full text] [doi: 10.1109/ICMLA.2013.89] [Medline: 28736774]
47. Bian J, Yoshigoe K, Hicks A, Yuan J, He Z, Xie M, et al. Mining Twitter to assess the public perception of the Internet of Things. PLoS One 2016;11(7):e0158450 [FREE Full text] [doi: 10.1371/journal.pone.0158450] [Medline: 27391760]

Abbreviations

Bi-LSTM: bidirectional long short-term memory
CNN: convolutional neural networks
KNN: k-nearest neighbors
SVM: support vector machines
CDC: Centers for Disease Control and Prevention
NLP: natural language processing

Edited by G Eysenbach; submitted 14.11.17; peer-reviewed by Z He, J Bian, N Limsopatham; comments to author 28.12.17; revised version received 01.04.18; accepted 10.05.18; published 09.07.18.

Please cite as:
Du J, Tang L, Xiang Y, Zhi D, Xu J, Song HY, Tao C
Public Perception Analysis of Tweets During the 2015 Measles Outbreak: Comparative Study Using Convolutional Neural Network Models
J Med Internet Res 2018;20(7):e236
URL: http://www.jmir.org/2018/7/e236/
doi:10.2196/jmir.9413
PMID:29986843

Original Paper

Rethinking the Meaning of Cloud Computing for Health Care: A Taxonomic Perspective and Future Research Directions

Fangjian Gao[1], MSc; Scott Thiebes[2], MSc; Ali Sunyaev[2], PhD

[1]Department of Information Systems, University of Cologne, Cologne, Germany

[2]Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany

Corresponding Author:
Ali Sunyaev, PhD
Department of Economics and Management
Karlsruhe Institute of Technology
Kaiserstraße 89
Karlsruhe,
Germany
Phone: 49 721 608 46037
Fax: 49 721 608 46581
Email: sunyaev@kit.edu

Abstract

Background: Cloud computing is an innovative paradigm that provides users with on-demand access to a shared pool of configurable computing resources such as servers, storage, and applications. Researchers claim that information technology (IT) services delivered via the cloud computing paradigm (ie, cloud computing services) provide major benefits for health care. However, due to a mismatch between our conceptual understanding of cloud computing for health care and the actual phenomenon in practice, the meaningful use of it for the health care industry cannot always be ensured. Although some studies have tried to conceptualize cloud computing or interpret this phenomenon for health care settings, they have mainly relied on its interpretation in a common context or have been heavily based on a general understanding of traditional health IT artifacts, leading to an insufficient or unspecific conceptual understanding of cloud computing for health care.

Objective: We aim to generate insights into the concept of cloud computing for health IT research. We propose a taxonomy that can serve as a fundamental mechanism for organizing knowledge about cloud computing services in health care organizations to gain a deepened, specific understanding of cloud computing in health care. With the taxonomy, we focus on conceptualizing the relevant properties of cloud computing for service delivery to health care organizations and highlighting their specific meanings for health care.

Methods: We employed a 2-stage approach in developing a taxonomy of cloud computing services for health care organizations. We conducted a structured literature review and 24 semistructured expert interviews in stage 1, drawing on data from theory and practice. In stage 2, we applied a systematic approach and relied on data from stage 1 to develop and evaluate the taxonomy using 14 iterations.

Results: Our taxonomy is composed of 8 dimensions and 28 characteristics that are relevant for cloud computing services in health care organizations. By applying the taxonomy to classify existing cloud computing services identified from the literature and expert interviews, which also serves as a part of the taxonomy, we identified 7 specificities of cloud computing in health care. These specificities challenge what we have learned about cloud computing in general contexts or in traditional health IT from the previous literature. The summarized specificities suggest research opportunities and exemplary research questions for future health IT research on cloud computing.

Conclusions: By relying on perspectives from a taxonomy for cloud computing services for health care organizations, this study provides a solid conceptual cornerstone for cloud computing in health care. Moreover, the identified specificities of cloud computing and the related future research opportunities will serve as a valuable roadmap to facilitate more research into cloud computing in health care.

(J Med Internet Res 2018;20(7):e10041) doi:10.2196/10041

KEYWORDS

cloud computing; taxonomy; health IT innovation

Introduction

Background and Objective

Cloud computing (CC) is an innovative paradigm that provides users with on-demand access to a shared pool of configurable computing resources such as servers, storage, and applications [1]. CC possesses unique features (ie, on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured services) that are argued to enhance traditional in-house health information technology (IT) approaches in health care organizations (eg, hospitals and clinics). Researchers claim that IT services delivered via the CC paradigm provide major benefits for health care, including improved flexibility in the use of IT resources [2], high availability of IT infrastructure to address ever-changing health IT demands [3], and low upfront investments and IT maintenance costs for the use of health IT [4]. Surprisingly, the benefits promised by using CC often do not hold in practice: it has, for example, been reported that the use of cloud computing services (CCSs) is tied to implementation and preparation activities that impede the flexibility of CC [5], the promised high availability of cloud-based IT infrastructures also cannot always be ensured (eg, sometimes the maximal attainable IT resources are strictly predefined) [6], and the use of CCSs is not guaranteed to yield the expected economic advantages for users in health care (eg, due to unexpected high upfront costs) [7,8]. There is therefore a mismatch between our conceptual understanding and the accepted meaning of CC for health care (ie, the value and/or consequences of using CC) in practice. Such a mismatch not only hampers the meaningful use of CC in the health care industry (ie, CC should provide constructive support) [9] but also could lead to countereffects for health care. As reported in a recent case, performance of an electronic health record system enabled by CC in a United Kingdom hospital diverged from initial expectations and led to countereffects, resulting in a £200 million (US $262 million) project failure and the hospital's inability to deliver key services on a large scale [10,11].

Although the topic of CC in health care has been widely discussed in the literature, existing publications mainly focus on development of single CC applications or platforms in health care [12-16] and development of security mechanisms for the use of CC [17-21]. Although some studies have tried to conceptualize CC or interpret this phenomenon for health care settings [4,22,23], they are heavily based on a general understanding of traditional health IT artifacts or mainly rely on the interpretation of CC in a common context, which leads to an insufficient or unspecific conceptual understanding of CC for health care. CC is an IT innovation for the health care industry that differs from traditional health IT approaches; in addition, when conceptualizing the topic of CC in health care, it is essential to seriously consider the health care context. The health care industry is markedly different from the commonly understood context and interpretation of CC [24]. Thus, this more general CC context is not necessarily adequate for health care. To this end, past research suggests that a nonspecific grasp of the CC concept in research and practice, irrespective of the intricacies of the health care sector, might be a major reason for why few successful implementations of CCSs in health care exist [25].

In this research, we rethink the meaning of CC for health care. By relying on existing CCSs in practice, we aim at generating insights into this phenomenon for health IT research. Our research focuses on the following research questions (RQs):

> RQ1: What are the relevant properties of CC for service delivery to health care?
>
> RQ2: What are the specific meanings of these properties for health care?

To address the research questions, we drew on data from a structured literature review and 24 expert interviews to develop a taxonomy of CCSs for health care organizations. Taxonomies are a form of classification [26] that are widely used to understand IT concepts in health care [27,28]. We expect to use this taxonomy to organize existing knowledge about CC in health care to fulfill our research purpose. In particular, we relied on the taxonomy to understand CC's key service delivery properties for health care organizations (RQ1) and thereby conceptualized CC for health care settings. By classifying 50 CCSs for health care organizations that we identified from both the literature and interviews using the taxonomy, we derived specificities of CC for health care (RQ2) that subverted and, therefore, challenged our understanding of CC in a common context or from a traditional health IT perspective. Our study conceptualizes CC specifically for health care. More importantly, we derived concrete research directions based on our conceptualization of CC to facilitate research on CC in health care.

Cloud Computing Knowledge in Health Care

CC is an innovation for health care organizations. In the health care industry, 3 types of innovations can be observed: (1) innovation focusing on the manner in which consumers access health care and fund the related services; (2) innovation applying technology to improve products, services, or care; and (3) innovation generating new business models [29]. CC is an innovation of applying (information) technology in health care organizations (type 2) that is in sharp contrast to traditional health IT approaches. CC provides 3 different service models—software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS)—all of which are Web-based [1]. CC can therefore deliver fundamental IT resources such as processing, storage (IaaS), and platforms together with programming languages, tools, and/or libraries that support users to develop and/or deploy software (PaaS). CC can also provide ready-to-use software applications (SaaS), which run on the cloud infrastructure, to health care organizations.

CC relies on different deployment models to provide IT services. First, in a public cloud, the infrastructure of CCSs is provided for open use by the general public. Second, the infrastructure

of a private or community cloud is provisioned for the exclusive use by a single organization or a specific group of organizations, respectively. Third, a hybrid cloud is a combination of 2 or more of the aforementioned deployment models. Whereas public clouds exist off the premises of cloud users, private and community clouds may exist on or off premises.

Our research aimed at organizing knowledge about CC and conceptualizing CC in health care. We employed the concept of knowledge about innovations by Rogers [30] as a means to interpret the knowledge about CC in health care and guide the taxonomy development. We chose it because Rogers' concept of knowledge is one of the few established concepts in research that can specify an IT artifact by observing it as an innovation, which is appropriate for CC as an innovation in health care. Moreover, Rogers' knowledge about innovations serves as a basic concept in his diffusion of innovations theory. Although we did not specifically address issues regarding CC's diffusion, we aimed for a specific understanding of an innovation (in health care), which is consistent with Rogers' ultimate purpose for this concept in the diffusion of innovations theory.

According to Rogers, 3 different types of knowledge are relevant for an insightful understanding of an innovation: (1) awareness knowledge comprises information about the existence of an innovation, (2) how-to knowledge describes how the innovation can be applied, and (3) principle knowledge explains the approach in which an innovation works. In this research, we targeted how-to and principle knowledge to understand the term knowledge. This is because most are aware of the term "cloud computing" [31]. Our research focused on the properties of CCSs that describe how CC can be used in health care organizations (how-to knowledge) and the ways in which CCSs support health care organizations (principle knowledge).

Methods

Overview

We employed a 2-stage approach to develop a taxonomy of CCSs for health care organizations. As illustrated in Figure 1, we conducted a structured literature review and 24 semistructured expert interviews in stage 1, drawing on data from theory and practice. In stage 2, we employed the views of how-to and principle knowledge, applied the method used by Nickerson et al [32], and developed a taxonomy of CCSs for health care organizations. The taxonomy development method integrates the evaluation of the taxonomy into its development process such that no further a posteriori evaluation of the taxonomy was required.

Literature Review

To obtain data for the development of our taxonomy, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses framework [33] and performed a review of the literature on CC in health care organizations. We searched literature databases to identify research articles addressing the topic of CC in health care organizations. Figure 2 presents a schematic of our approach, which includes the literature databases and the search string employed. It must be emphasized that we iteratively developed our search string. We tested broader keywords (eg, "eHealth," "health IT") but decided to employ more specific keywords that target health care organizations for the final search string because our taxonomy specifically focused on health care organizations. Moreover, we found that the broader keywords did not result in many additional relevant articles but increased noise, which diminished the quality of the literature review. We performed keyword, title, and abstract searches and ultimately full-text reviews. Next, 2 researchers independently screened the identified articles. The articles were first screened using keywords, titles, and abstracts and then using the full texts. We excluded articles that were not published within the last 10 years (not up to date: the term CC was not readily used until 2007), not in English, not peer-reviewed, or did not address the topic of CC in health care organizations (off-topic). A total of 66 articles remained after the screening.

Figure 1. Research methods overview. Asterisk refers to taxonomy evaluation by means of the ending conditions. CC: cloud computing, CCS: cloud computing service.

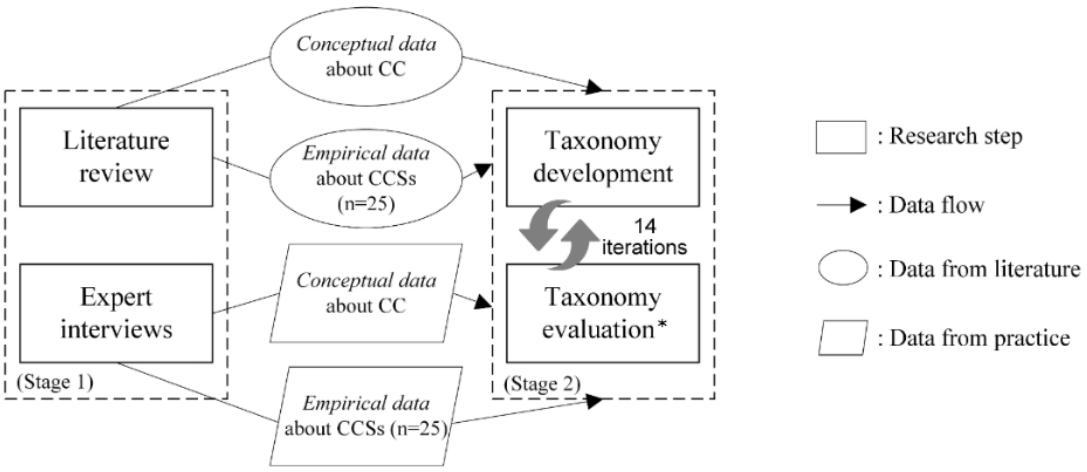
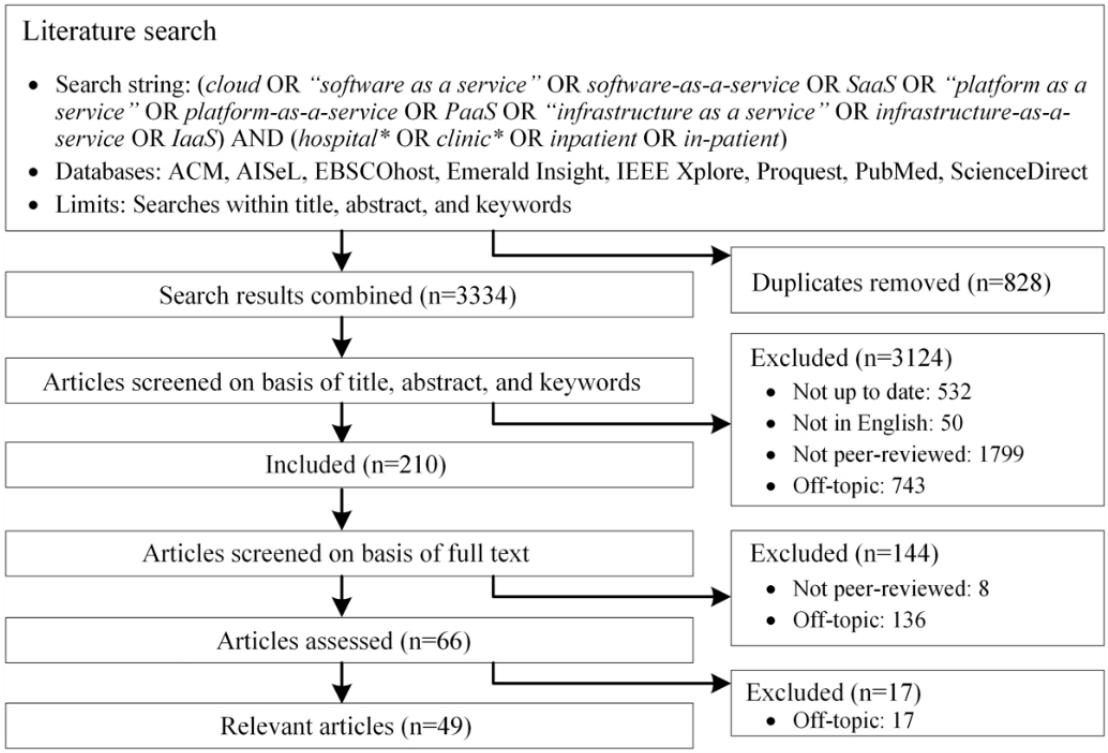


Figure 2. Flow diagram of inclusion/exclusion and literature analysis.



Once the screening was complete, we analyzed the remaining articles and identified 17 additional articles that were off-topic but could not have been excluded without an in-depth full-text assessment. This process resulted in a final sample of 49 eligible articles that were assessed in detail. With the assessment, we aimed to understand the concept of CC in health care organization contexts from a research perspective. Moreover, we attempted to identify concrete CCSs for health care organizations in addition to their characteristics from the literature. Accordingly, we classified the literature into 2 categories: conceptual and empirical. The conceptual category covered articles providing general conceptual statements about CC in health care and articles proposing CCSs that have not been deployed in practice. The empirical category contained articles describing concrete CCSs for health care organizations. This occurred because the applied taxonomy development method employed both a deductive approach (development based on data from the conceptual category) and an inductive approach (development by observing objects that need to be classified, namely, data from the empirical category) [32]. Of the 49 eligible articles, 24 were classified as conceptual and 25 as empirical. Articles that describe general features of CC and apply them to concrete CCSs were classified as special cases of the empirical category. Two researchers separately analyzed the articles. Each relevant statement was extracted and converted into 1 or more pieces of code representing a property of CCSs for health care organizations. Codes created by both researchers were compared and aggregated resulting in a master list containing codes encapsulating the properties of CCSs. The master list covers codes from both the conceptual (ie, general conceptual understanding of CC) and empirical categories (ie, concrete CCSs and their properties). It must be emphasized that 25 concrete CCSs for health care organizations were identified from the literature. A description of these CCSs can be found in Multimedia Appendix 1.

Expert Interviews

To gather knowledge that could inform the development of the taxonomy from practice, we conducted 24 semistructured expert interviews, as listed in Table 1. We applied a purposeful sampling strategy that focused on selecting individuals who are especially knowledgeable about a phenomenon of interest to recruit interviewees [34]. We included only experts who were engaged in IT activities in health care organizations and who had used, provided, or knew about concrete CCSs for health care organizations. After 24 interviews, we reached data saturation and stopped recruiting additional interviewees. The first 12 interviewees listed in Table 1 focus on the Chinese health care cloud market, and the rest focus on the German market. We selected these countries because they are the main cloud players in Asia and Western Europe, which are among the regions with the highest market share in the overall [35] and the health care cloud markets [36]. Moreover, the cloud markets in China and Germany are complementary to each other: whereas CCSs for health care organizations in Germany are restricted to European cloud providers due to data protection regulations by the European Union, CCSs in China rely on large health IT players (eg, IBM, Cisco, and Microsoft) mainly from the United States supplemented by Chinese domestic providers [37]. Thus, we were able to gain insights into knowledge about CC in health care from a wide spectrum of practices. The interviewees came from 18 different organizations and had an average of 15 years of work experience.

Table 1. Overview of interviewees.

ID	Job title	Experience in health IT[a] (years)	Work organization
i01	Chief information officer	8	General hospital in China
i02	Chief of information center	18	General hospital in China
i03	Project manager	12	International health IT provider
i04	Staff of new media department	6	Specialized hospital in China
i05	Chief of IT department	15	District clinic in China
i06	Chief executive officer	16	Chinese health IT provider for dental clinics
i07	Senior IT staff	12	General hospital in China
i08	IT supervisor	17	Chinese governmental organization for the strategic development of public hospitals
i09	Chief of information center	11	General hospital in China
i10	Senior IT staff	9	General hospital in China
i11	Vice director	12	District hospital in China
i12	Head of IT	6	General hospital in China
i13	Chief marketing officer	33	Health IT provider for the German market
i14	Staff of research and development department	30	Health IT provider for the German market
i15	Head of IT applications	20	University clinic in Germany
i16	Technology officer	10	Health IT provider for the German market
i17	Head of IT development	6	German local health IT provider
i18	Health IT developer	6	German local health IT provider
i19	Senior manager	19	German local health IT provider
i20	Head of IT	17	University clinic in Germany
i21	IT staff	10	University clinic in Germany
i22	IT team leader	19	University clinic in Germany
i23	Chief information officer	12	District hospital in Germany
i24	Head of IT infrastructure	31	University clinic in Germany

[a]IT: information technology.

Our interview guide was structured into 3 topics, as shown in Multimedia Appendix 2. Topic 1 addressed the interviewee's organization, work activities, and professional experience. Topic 2 focused on the interviewee's (conceptual) understanding of CC in health care. In topic 3, interviewees were asked to enumerate and describe all concrete CCSs in health care organizations with which they were familiar. The interviews lasted between 30 and 90 minutes, with an average of 51.33 minutes. All interviews were audio recorded and transcribed afterwards.

Two researchers separately analyzed the transcripts. For the same reasons as in the literature analysis, the interview analysis focused on not only the conceptual understanding of CC in health care but also concrete examples of CCSs, including their properties. Thus, we classified the interview data obtained from topic 2 of the interview guide in the conceptual category, whereas the interview data obtained from topic 3 fell into the empirical category. Both researchers employed the same coding technique used in the literature analysis to analyze the interview data. Consequently, we obtained a list of codes representing a conceptual view of CC in health care for the conceptual category and a list of codes representing properties of concrete CCSs in health care organizations for the empirical category. In total, 25 CCSs for health care organizations were identified from the interviews, which are presented together with the 25 CCSs identified from the literature in Multimedia Appendix 1.

Taxonomy Development

For the taxonomy development, we chose the method proposed by Nickerson et al [32], which provides a systematic taxonomy development approach for IT objects and is well acknowledged in the domain of health IT [38,39]. According to Nickerson et al [32], a taxonomy is a set of dimensions in which each dimension consists of more than 1 characteristic. In taxonomy development, several iterations are used to determine dimensions and characteristics. After each iteration, predefined ending conditions are employed to evaluate the taxonomy: if not all ending conditions can be fulfilled, the taxonomy development continues with the next iteration. In each iteration, researchers can choose between an inductive and deductive approach. A deductive approach is based on theoretical knowledge about the objects that need to be classified; an inductive approach is

based on observing and analyzing a sample of the objects. For the deductive approach, we applied all data about CC from the conceptual category (see Figure 1). For the inductive approach, we employed data from the empirical category for all 50 identified CCSs in health care organizations.

Before developing a taxonomy, researchers must define a meta-characteristic and ending conditions. The meta-characteristic guides the choice of dimensions and characteristics in the taxonomy. As a result, each dimension or characteristic of the taxonomy is a logical consequence of the meta-characteristic. Our taxonomy builds on 2 relevant knowledge types of CCSs to define the meta-characteristic: how-to and principle knowledge. We defined "service delivery properties of CCSs for health care organizations" as our meta-characteristic that covers how CCSs can be used by health care organizations (how-to knowledge) and describes the approaches in which CCSs support them (principle knowledge). Both knowledge types serve as the conceptual orientation of the taxonomy as a whole. For the ending conditions, we adopted all of the objective and subjective ending conditions from Nickerson et al [32]. The subjective ending conditions also serve as criteria to evaluate the sufficiency of the taxonomy.

For each iteration, we randomly chose a developmental approach (ie, inductive or deductive). Based on the chosen approach, we randomly selected data from our data pool accordingly (ie, understanding of CC from the conceptual category for a deductive approach and concrete CCSs and their properties from the empirical category for an inductive approach). The amount of data was adjusted such that each iteration could be performed in a reasonable time frame (45 to 60 minutes).

For an iteration using the deductive approach, we first examined codes about CC to identify and summarize new characteristics and/or dimensions. We determined whether each potential new characteristic or dimension derived from a code could be considered a logical consequence of the meta-characteristic and whether there was a concrete CCS in our empirical category that could be classified into this characteristic/dimension. If both criteria were fulfilled, the new characteristic/dimension was added to the existing taxonomy. For an iteration using the inductive approach, we first examined and compared the properties of the selected CCSs from the empirical category. We attempted to derive common characteristics of the chosen CCSs by comparing their codes. If the identified characteristics were new, we attempted to assign them to existing dimensions (as characteristics) if possible. Otherwise, we grouped the characteristics, inspected their conformity with the meta-characteristic, and defined them as new dimensions for the taxonomy, if necessary. After each iteration, we applied the predefined ending conditions to evaluate our taxonomy. For an inductive approach, we additionally classified all CCSs that were analyzed using the (preliminary) taxonomy, as required by Nickerson et al [32]. After 14 iterations, we met all ending conditions and thus stopped the taxonomy development. Multimedia Appendix 3 summarizes these iterations and the data we applied to each. Because all identified CCSs for health care organizations (n=50) were analyzed in our research (ie, an objective ending condition), these CCSs were classified by the taxonomy. The final classification result serves as a part of the taxonomy.

Results

Dimensions and Characteristics

Our taxonomy of CCSs for health care organizations is composed of 8 dimensions and 28 characteristics (see Table 2 for overview). The first 4 dimensions (service form, deployment model, targeted cloud advantage, and timeliness) represent principle knowledge, which is related to the inherent mechanisms and principles of a CCS and describes the approaches in which CC supports health care organizations. The remaining 4 dimensions address concrete methods to implement (ie, how to use) CCSs for health care and represent how-to knowledge.

The service form and deployment model dimensions are consistent with the service and deployment models of CC, respectively [1]. They clarify the most basic operational principles of CCSs for health care organizations, which relate to principle knowledge. The dimension service form contains 3 characteristics: infrastructure, platform, and software, which refer to IaaS, PaaS, and SaaS of CC, respectively. The deployment model dimension indicates whether CCSs are deployed using a public, community, or private cloud. Because a hybrid cloud is, by definition, composed of 2 or more of the aforementioned deployment models, we do not define hybrid as an independent characteristic of the deployment model. Instead, our taxonomy represents a CCS with a hybrid deployment model by using 2 or more of the characteristics defined above.

The targeted cloud advantage dimension describes the concrete cloud properties from which a health care organization can benefit. This dimension highlights the effects of using CCSs and is also considered a type of principle knowledge. Scalability refers to the advantage of a CCS that extends its IT resources (eg, storage, processing, and memory) to overcome a health care organization's IT resource scarcity or support resource-intensive tasks. Elasticity represents a CCS's capability to dynamically allocate available resources based on users' demands and thus optimize resource use for all users. Ubiquity indicates that users can access the CCS from any location. Cost efficiency emphasizes the cost advantage brought by CCSs. Shareability refers to the ability of CCSs to enable the efficient exchange and sharing of data between different users, whereas interoperability denotes the ability of a CCS to smoothly integrate and operate with disparate systems and machines. Security allows health care organizations to take advantage of cloud providers' advanced data security mechanisms or technologies.

Timeliness assesses how quickly CC is able to deliver services and related data to health care organizations (real time vs not real time) and thus relates to principle knowledge. We define a CCS as real time if it is ready to process or transfer data at any time, such that the computational results and requested data are immediately available.

Table 2. Taxonomy of cloud computing services for health care organizations.

Dimension	Characteristics
Principle knowledge	
Service form	Software, platform, infrastructure
Deployment model	Public, private, community
Targeted cloud advantage	Scalability, elasticity, ubiquity, cost efficiency, shareability, interoperability, security
Timeliness	Real time, not real time
How-to knowledge	
Supported task	Clinical, administrative, strategy, research
User	Patient, medical staff, family member
Service delivery device	Independent, adapted, specialized
Patient data involvement	Internal, external, no involvement

The supported task dimension specifies the areas in which health care organizations use CCSs. This dimension highlights the manner in which CC supports health care and is deemed a type of how-to knowledge. Supported task includes 4 characteristics: clinical, administrative, strategic, and research. Clinical refers to medical activities in health care organizations that are directly associated with patient diagnosis and treatment. Administrative denotes management or support tasks in health care organizations, such as patient registration, admission, and discharge. Strategic represents tasks performed by management teams in health care organizations, such as strategic planning decisions, human resources management, and performance evaluations. Research represents all activities that are related to medical research.

The user dimension relates to how-to knowledge and aggregates the possible user types of CCSs. This dimension differentiates between a patient who receives medical treatment at a health care organization, the medical staff (health care professionals as well as administrators), and the family members of the patient.

Service delivery device refers to how-to knowledge because this dimension represents the types of client devices used to access the CCS. A CCS with an independent characteristic allows users to access services using any computer or mobile device. Adapted specifies that a CCS is compatible with different types of devices but operates more efficiently on a certain group of devices (eg, mobile phones or tablets) via technical adaptation to those devices (eg, developing specialized applications for tablets or compressing data to accelerate data transfer for mobile phones). Specialized represents those CCSs that can be accessed by only 1 or several designated groups of devices, such as authorized tablet computers, workstations in health care organizations, or specific medical devices.

Finally, the patient data involvement dimension, which also relates to how-to knowledge, explains how patient-related data are used to deliver services. Internal indicates that a CCS uses patient data that are internally available to the health care organization for IT service delivery. External refers to a situation in which a CCS uses patient data collected from external sources, such as outside medical professionals or the patients themselves. No involvement indicates that a CCS does not have access to patient data and thus does not use such data in IT service delivery.

Classification and Evaluation

After completing all taxonomy development iterations, we classified all 50 CCSs that we identified during stage 1. Multimedia Appendix 4 presents the final classification results. In this section, we provide an example of how our taxonomy can be used to classify CCSs for health care organizations. This example examines a hospital decision support system for bed-patient assignments (see C22, Multimedia Appendix 1). Because this CCS addresses patient administration and assists hospital leadership in measuring and benchmarking hospital operations, it supports both administrative and strategic tasks. The CCS is delivered in the form of a software application and is hosted in a public cloud environment. The targeted cloud advantage is scalability because the hospital benefits from CC's computing resources to analyze large quantities of data based on complex mathematical models. The CCS does not operate in real time (not real time). It is used by medical staff and is not device-specific (independent). Finally, the patient data processed by the CCS are internal.

Our taxonomy fulfills all predefined ending conditions after 14 development iterations. In particular, the fulfillment of 5 subjective ending conditions indicates high sufficiency of the taxonomy. We summarized these subjective ending conditions and provide a justification for the fulfillment of each condition in Multimedia Appendix 5. Notably, the subjective ending conditions describe the essential features of the derived taxonomy.

Discussion

Principal Findings

Specific Meanings of Cloud Computing for Health Care and Research Opportunities

By observing the taxonomy, which includes the classification results of CCSs for health care organizations, we obtained specific implications of CCSs for health care.

Table 3. Specificities of cloud computing for health care.

Number	Specificity	Previous understanding	Type
1	CC[a] relies on SaaS[b]	PaaS[c] and IaaS[d] in general are as relevant as SaaS	Type 1[e]
2	CC increases data security and interoperability	Low data security and interoperability as CC's downside	Type 1
3	If any, CC only brings economic benefits in the long term	Reduced costs by using CC in general	Type 1
4	CC focuses on clinical tasks	Health IT[f] traditionally supports more management areas	Type 2[g]
5	CC supports patient-centeredness	Health IT products are traditionally heavily physician-centered	Type 2
6	CC increases service mobility and flexibility	Health IT traditionally suffers from inflexible service access	Type 2
7	CC facilitates collaboration in clinical areas	Insufficient capabilities of traditional health IT to support collaboration	Type 2

[a]CC: cloud computing.

[b]SaaS: software as a service.

[c]PaaS: platofrm as a service.

[d]IaaS: infrastructure as a service.

[e]The specificity challenges what we have learned about CC in a general context.

[f]IT: information technology.

[g]The specificity challenges what we have learned about traditional health IT.

As demonstrated in Table 3, these implications offer 2 types of challenges to our previous understanding of CC in health care: they challenge what we have learned about CC in a general context (type 1) and in published traditional health IT studies (type 2). We employed the term "specificities" to summarize these implications, thereby highlighting the specific meanings of CC for health care. More importantly, as shown in Figure 3, the summarized specificities suggest research opportunities with exemplary research questions, facilitating future research about this relevant phenomenon in health IT.

Specificity 1: Cloud Computing in Health Care Relies on Software as a Service

Previous studies show that in a common context, PaaS and IaaS are as relevant as SaaS in the cloud market [40]; however, this result is challenged by CC in the context of health care (type 1). We found that 92% (46/50) of the CSSs deliver services in the form of SaaS (dimension service form). The identified research articles and the interviewees even applied the term "X as a service," such as "hospital information system as a service" [41] or "documentation as a service" (i17), to emphasize the importance of such CCSs, although by their nature they belong to SaaS. This is possibly because health care organizations expect to exploit the advantages of SaaS to the greatest extent and in a timely manner.

> *For hospitals, cloud almost only means software as a service because many hospitals want to use (them as) off-the-shelf products. ...SaaS products that support medical areas are especially welcome because hospitals always expect to get immediate improvement from the cloud in their core business.* [Interviewee i03]

The lack of PaaS and IaaS in health care organizations indicates an insufficient state of CC in health care, which was confirmed by several interviewees (i07-i08, i10, i17-i19). For PaaS, our taxonomy shows only one CCS (C06), although several interviewees noted the urgent need for industry-specific PaaS.

> *We want to develop our own SaaS, but there is just no specific PaaS for health care organizations. General PaaS are not enough.* [Interviewee i07]

The need for PaaS in health care is not only because PaaS in general provides ready-to-use technical support for programmers but also because it has the potential to provide solutions to effectively fulfill industry-specific IT requirements. This is, for example, explained by an interviewee who was involved in developing a CCS for a hospital.

> *There were so many complex things we had to consider for hospitals. We kept wasting time on unnecessary meetings to find technical solutions. I dreamt of having a PaaS that could support us. ...Of course, there is more. ...Compliance is also a main topic. Hospitals ask over and over again whether our software is compliant with this or that. ...Example HIPAA: If the PaaS we use is compliant with HIPAA, then we can tell them: Yes, our software is HIPAA-compliant.* [Interviewee i17]

Further industry-specific IT requirements that can potentially be supported by a health care PaaS—constant demand on cutting-edge technologies, high health IT agility (to meet changing medical requirements), the need for different domain-specific medical data structures, and support for industrial joint implementation activities (eg, between government and hospital)—were also mentioned by the interviewees.

For IaaS, previous research studies [42] and our interviewees both emphasized the strategic meaning (i08) of IT infrastructure (ie, critical information infrastructure) for the health care industry and consequently the extremely high importance of IaaS (i20) for health care organizations. We identified only a limited number of IaaS (n=3) used for general administration of health care organizations (C28, C37) or data storage (C38), which hardly fulfills all health care organization IT infrastructure requirements.

Figure 3. Research opportunities for cloud computing in health care. CC: cloud computing, CCS: cloud computing service, IaaS: infrastructure as a service, IT: information technology, PaaS: platform as a service.

Future research could focus on exploring the lack of PaaS and IaaS for health care. As revealed by our interview data, there is a particular need for research studies that systematically investigate specific requirements for health care that cannot be covered by PaaS and IaaS in a common context and thus a need to design and develop industry-specific PaaS and IaaS.

Specificity 2: Cloud Computing Brings More Data Security and Interoperability to Health Care

Previous studies have raised concerns about security and privacy as the Achilles heel of CC [43], which are main barriers for the adoption of health IT artifacts [44,45]. These concerns might be more severe for public clouds, whose infrastructures are accessible by many different users [46]. However, the dimension deployment model indicates that more than half of the investigated CCSs are based on public clouds, especially given that almost all of these CCSs involved patient data (dimension: patient data involvement) that were sensitive and entailed security or privacy issues. To this end, providing a high level of data security was regarded as a targeted cloud advantage in 10 of the identified CCSs, of which 6 were deployed on public clouds. This challenges our understanding of CC in a general context (type 1). Additionally, interoperability may also impede the adoption of CC in a general context [47]. For health care, however, our taxonomy demonstrates that increased interoperability is a benefit of CC. Security and interoperability are traditionally the most intractable challenges in health IT, and industry standards concerning IT security and interoperability in health care are evolving [9]. Cloud providers can devote resources to the implementation of industry standards or best practices that many hospitals cannot afford [4]. CC can thereby address security and interoperability issues in a more effective manner, which was confirmed by the interviewed experts (i03-i04, i06-i07, i10, i13-i14, i16-i18, i21).

> *CC is safe. The problem is how to make people believe that.* [Interviewee i13]

> *Data security, interoperability...these are pluses. Speaking of data security, using paper is also not safe, if you insist on saying a cloud is not safe.* [Interviewee i21]

As highlighted in Figure 3, future research could investigate the role of security and interoperability in cloud adoption studies and focus on the adopter's awareness or perception of increased data security and interoperability from CC in health care settings. Moreover, researchers could focus on exploring the factors (such as security and interoperability) that have industry-specific impacts on cloud adoption in health care, in contrast to a general context.

Specificity 3: Cloud Computing Brings Economic Benefits to Health Care Organizations, if Any, Only in the Long Term

It is surprising that CC offered economic advantages (cost efficiency) for only 11 of the 50 CCSs. In a general context, the use of CC is heavily motivated by short-term economic interests [48]. Research relying on this general understanding of CC claimed the low costs were the principle advantage of CC in health care [4]. Our research challenges the understanding of CC in a general context (type 1) by revealing that when using CCSs, many health care organizations frequently must transfer large volumes of data to and from the cloud (eg, medical images [49]). This can cause data transfer bottlenecks due to the obsolete (network) infrastructures currently in place at many health care organizations—a typical industry-specific IT issue (i02, i08, i15). Thus, CC might still require significant short-term investments in health care organizations' network resources, internet bandwidth, or other relevant infrastructures. It is therefore not surprising that the interviewees were not convinced of the potential financial advantages of using CC in health care (i01-i05, i07, i10, i17). They (i01-i02, i10) even noted that additional expenses for CC, such as consulting fees, could increase health care organizations' expenses. However, our interviewees reported that in the long term, CC will reduce their general IT maintenance work (i02, i24) and help them avoid possible IT reinvestments (i22). Future research could therefore focus on (re)examining and explaining the economic results of using CCSs in health care organizations. Moreover, researchers could focus on CC business processes or investment strategies in health care settings that enhance the short-term benefits for health care organizations.

Specificity 4: Cloud Computing Mainly Focuses on Clinical Tasks (by Leveraging High Scalability)

We recognize that most of the identified CCSs (36 of 50) support clinical tasks in health care organizations (dimension: supported task). This observation challenges previous studies about traditional health IT (type 2), which have concluded that health care organizations primarily focus on the use of IT applications for administrative, strategic, or financial functions rather than clinical activities [50]. These findings reflect an urgent need to use CC to remedy the deficiencies of traditional health IT in the context of health care organizations' clinical activities, as revealed by our literature review [51].

> *In clinical practice, even ordinary data analysis occasionally overwhelms traditional health IT with large volumes of data and complex analytical algorithms.* [Interviewee i16]

CC can address this problem with highly scalable IT resources and is therefore considered a "powerful weapon for IT tasks in the clinical area" (Interviewee i03).

This viewpoint is supported by our taxonomy, as more than 70% (23/32) of the CCSs possessed high scalability as one of their advantages (dimension: targeted cloud advantage), with a focus on clinical areas. For research opportunities, we suggest researchers concentrate on CC that supports research tasks in health care because both the literature [52] and our interviewees (eg, i18) reveal that research activities in health care depend even more on highly scalable IT resources to address large amounts of data, which is currently managed only in a small number of identified CCSs (n=6).

Specificity 5: Cloud Computing Supports Patient-Centeredness

A conservative but still well-recognized view of health IT is that medical staff are the main users of health IT applications [53,54], and many existing health IT applications are heavily

physician-centered. However, the evidence from our taxonomy challenges this view (type 2) and implies a high potential of CC to realize patient-centeredness—a promising future direction for health IT [55]. Regarding the user dimension, we noticed that 8 identified CCSs included patients as their users, which is a premise of patient-centered health IT services. Among them, 7 CCSs were patient-centered (C05, C07, C10, C26, C29, C32, C34), as they possessed 3 essential attributes of patient-centered health IT: patient-focused, patient-active, and patient-empowered [56]. Additionally, several interviewees (i02, i07-i08, i11) noted that CC innovatively involves patient family members to realize patient-centeredness, as did 2 identified CCSs (C26, C29). An interviewee, whose hospital deploys a medical appointment CCS for patients, had this to say:

> *Seniors, the disabled, or someone who doesn't like technologies also needs to use appointment services, so we decided to involve their relatives. ...Although we have to have more users and processes now, I believe CC can offer the necessary computer resources. It's a good thing, and I think this might be a reason to have more CCSs.* [Interviewee i02]

> *We even have some patients who don't use the Internet at all. Their children could help them...only in this way can we ensure that each patient truly benefits from our services.* [Interviewee i08]

Despite the potential of CC to support patient-centeredness, only a limited number of patient-centered CCSs were identified in this study. Future research could therefore focus on examining how CC supports patient-centeredness and on designing further CCSs that support it.

Specificity 6: Cloud Computing Increases Service Mobility and Flexibility

We found that 42% (21/50) of the identified CCSs adapt themselves to or are specialized for certain devices for service delivery (dimension: service delivery device). For CCSs that support clinical tasks, this rate is even higher (16/36, 45%). In general, a barrier impeding the use of health IT is the alteration of users' traditional workflow paradigm [57]. For health IT that supports clinical functions, physicians who are forced to adapt health care delivery processes to technologies are often unwilling to use it. Our taxonomy reveals that almost 80% (16/21) of the CCSs that were adapted to user devices, such as mobile phones and tablet PCs or other specialized medical devices, targeted service ubiquity (dimension: targeted cloud advantage) and thus the mobility and flexibility of IT service delivery (type 2). Existing health IT research concluded that these devices are inherently subjected to limited computing capacity and are criticized as unsuitable for complex tasks, such as clinical work [58]. However, our research shows that more than one-third (8/21) of the CCSs that were adapted to user devices enjoyed the benefit of resource scalability (dimension: targeted cloud advantage). Thus, as emphasized by our interviewees, CC can effectively "offset the [traditional] limitations of mobile devices or other small devices. It can increase the use of innovative devices in health care" (Interviewee i07). Future research could explore how CC overcomes the limitations of mobile or small devices in health care, which is a relevant but underinvestigated topic in health IT [58].

Specificity 7: Cloud Computing Facilitates Collaboration in Clinical Settings.

Our taxonomy demonstrates that most of the CCSs (46/50) involved the use of patient data (dimension: patient data involvement). One major expected purpose of involving patient data in health IT is to employ the data as a means to link users or systems in different clinical areas and thereby facilitate their collaboration [59]. However, research generally highlights a lack of sufficient health IT applications that support collaboration [60]. Our taxonomy challenges this (type 2) and reveals that CC has the potential to address this issue, as 21 of the 46 CCSs (that involve patient data and support clinical areas) possessed shareability or interoperability as an advantage (dimension: targeted cloud advantage) and had improved collaboration between users or systems as one of their main purposes. However, these CCSs are not without limitations. Only a small fraction of these CCSs (6/21) involved patient data from external sources (dimension: patient data involved). Including patient data from different sources is the basis of collaboration in clinical activities [51]. Our interviewees (i02, i05, i08, i11, i15) noted that including patient data from external sources (eg, external medical professionals or patients themselves) is relevant for improving collaboration in clinical processes because "no hospitals can depend only on themselves. They need continual cooperation with, at least, patients" (Interviewee i02).

The interviewees remarked that CCSs in health care organizations that have a collaboration purpose mostly focus on internal data exchanges (which was also revealed by our taxonomy), although they believed that CC has the potential to also facilitate collaboration with external parties. The timeliness dimension is another indicator for collaboration because it addresses how intensively data exchanges occur. However, for the 21 CCSs that supported clinical areas and possessed the shareability or interoperability characteristics, we found that only 8 enabled real-time data exchanges. Real time is crucial for effective data exchanges and the resulting collaboration in clinical processes (i05-i06, i08, i11, i18).

> *Collaboration [based on data exchanges] should not only take place but also in a real-time manner. A delay of important data for even a few minutes could be fatal for clinical activities.* [Interviewee i08]

Future research should therefore strive to improve CCSs for collaboration in clinical activities due to the currently (still) insufficient state of CCSs (as well as general health IT [51,60]) for supporting collaboration. Moreover, researchers could also investigate how CC supports collaboration in areas other than clinical settings in health care.

Contributions

For health IT research, our contributions are threefold. First, we suggest a taxonomy that structures the knowledge of CCSs (ie, CCS properties) for health care organizations. In particular, our taxonomy targets principle and how-to knowledge to

systematically conceptualize the concept of CC for health care settings. Unlike previous research that heavily relied on CC literature from common contexts or on traditional understandings of health IT, our study analyzed CC's industry-specific properties not only from the health IT literature but also from practice. Thus, the derived dimensions and characteristics of the taxonomy highlight the aspects of CC that are most relevant to health care. We thereby contribute to closing the gap between an insufficient conceptual understanding of CC and the actual phenomenon in practice for health care. Second, our taxonomy suggests 7 specificities that subvert and thus challenge our previous understanding of CC in a general context or of traditional health IT. These specificities advance the understanding of CC in health care. Third, we derived concrete research opportunities for health IT (see Multimedia Appendix 6 for a summary). As presented at the beginning, health IT researchers have been interested in the development of single CC applications or data security topics. For both topics, we provide suggestions that guide future research (eg, to focus on developing CCSs that enable collaboration in health care) or even create new opportunities and directions (eg, to focus on inherently increased, instead of decreased, IT security in health care by using CC). In addition, we noticed that research topics on CC are by nature broad and diverse, which should not be limited to the development of CC applications and IT security, as in current health care settings, but can include more areas such as its business perspective [61,62], its adoption (by organizations) [63,64], user awareness and acceptance [65,66], and its certification [67-69]. The proposed research directions in this study are a step toward facilitating research on CC in health care settings.

For health IT practice, the derived taxonomy can be applied to investigate CCSs for health care organizations on 2 different levels. On a macro level, the classification of available CCSs in a certain health IT market using the taxonomy can serve as an indicator of the current state of these CCSs. Cloud providers or policy makers could, for example, suggest new CCSs that address possible market gaps (eg, PaaS for hospitals). On a micro level, health care organizations could apply the taxonomy to understand an individual CCS. In particular, by combining the characteristics from the dimensions that a CCS possesses, health care organizations could specify each CCS's profile as demonstrated, for example, by the hospital decision support system for bed-patient assignments, as referred to in the Results section. By finding matches as well as mismatches between the CCS's profile and their own organizational needs, health care organizations could screen and identify CCSs that would be useful to them and thereby increase the meaningful use of CC.

Limitations and Conclusions

A main limitation of this research is that our data focused on health care organizations that are hospitals and clinics, as implied by the literature review search string and by the interview questions. This is because hospitals and clinics are not only the backbone of the health care industry [70] but also representative IT consumers in health care [71]. We therefore expected that a taxonomy derived from hospitals and clinics would provide more generally valid insights into CC for health care settings. Research that focuses on CC in more specific health care settings (eg, nursing homes) could employ our taxonomy as a starting point. We suggest that such research use the proposed dimensions and characteristics as a checklist to investigate CC. If required, adjustments along the taxonomy's dimensions and/or characteristics can be easily carried out [32], resulting in more specific taxonomies that are useful for certain health care settings. Future research should also broaden the perspective on the topic of CC to cover further health care settings by using, for example, more general search strings for literature reviews (eg, including terms such as "health IT" and "eHealth") or by designing interview topics that cover CCSs in other health care areas.

Our work relied on data from 24 expert interviews, which does not necessarily guarantee that all CCSs for health care organizations from practice were discovered. However, the selection of our interviewees ensured a wide spectrum of knowledge about CC in health care in Asia, Western Europe, and the United States, which represent the main CC health care markets. Future research could also include niche CC markets to further verify and improve our taxonomy.

Although the term "cloud computing" has existed since 2007, the phenomenon of CC in health care remains in its infancy and calls for research on this phenomenon have emerged [4,25]. By relying on perspectives from a taxonomy for CCSs for health care organizations, we provide a solid conceptual cornerstone for research about CC in health care; moreover, the suggested specificities of CC for health care and the related future research opportunities will serve as a valuable roadmap.

Acknowledgments

We acknowledge support by Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of Karlsruhe Institute of Technology. We acknowledge support by the Multi-Disciplinary Identification of Lineage-Specific Signaling Dependencies in Cancer research project (FKZ 01ZX1406/01ZX1615), funded by the German Federal Ministry of Education and Research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of identified cloud computing services.

[PDF File (Adobe PDF File), 468KB - jmir_v20i7e10041_app1.pdf]

Multimedia Appendix 2

Overview of interview questions.

[PDF File (Adobe PDF File), 415KB - jmir_v20i7e10041_app2.pdf]

Multimedia Appendix 3

Taxonomy development iterations.

[PDF File (Adobe PDF File), 31KB - jmir_v20i7e10041_app3.pdf]

Multimedia Appendix 4

Taxonomy of cloud computing services for health care organizations.

[PDF File (Adobe PDF File), 595KB - jmir_v20i7e10041_app4.pdf]

Multimedia Appendix 5

Taxonomy's fulfillment of the subjective ending conditions.

[PDF File (Adobe PDF File), 427KB - jmir_v20i7e10041_app5.pdf]

Multimedia Appendix 6

Future research directions.

[PDF File (Adobe PDF File), 480KB - jmir_v20i7e10041_app6.pdf]

References

1. Mell P, Grance T. NIST definition of cloud computing.: National Institute of Standards and Technology; 2011. URL: http://dx.doi.org/10.6028/NIST.SP.800-145 [accessed 2018-06-27] [WebCite Cache ID 6wspXxvR5]
2. Paul M, Das A. Provisioning of healthcare service in cloud. In: Mishra DK, Azar AT, Joshi A, editors. Information and Communication Technology. Singapore: Springer; 2018:259-268.
3. Quwaider M, Jararweh Y. Multi-tier cloud infrastructure support for reliable global health awareness system. Simulation Modelling Pract Theory 2016 Sep;67:44-58. [doi: 10.1016/j.simpat.2016.06.005]
4. Kuo AM. Opportunities and challenges of cloud computing to improve health care services. J Med Internet Res 2011;13(3):e67 [FREE Full text] [doi: 10.2196/jmir.1867] [Medline: 21937354]
5. Sultan N. Making use of cloud computing for healthcare provision: opportunities and challenges. Int J Inform Manag 2014 Apr;34(2):177-184. [doi: 10.1016/j.ijinfomgt.2013.12.011]
6. The key lessons of the Healthcare.gov fiasco.: ZapThink; 2013. URL: https://www.cloudcomputing-news.net/news/2013/nov/01/key-lessons-of-the-healthcaregov-fiasco/ [accessed 2018-06-27] [WebCite Cache ID 6wstOkEUl]
7. Knaus J, Hieke S, Binder H, Schwarzer G. Costs of cloud computing for a biometry department. A case study. Methods Inf Med 2013;52(1):72-79. [doi: 10.3414/ME11-02-0048] [Medline: 23188548]
8. Miah SJ, Hasan J, Gammack JG. On-cloud healthcare clinic: an e-health consultancy approach for remote communities in a developing country. Telematics and Informatics 2017 Feb;34(1):311-322. [doi: 10.1016/j.tele.2016.05.008]
9. Nelson R, Staggers N. Health Informatics: An Interprofessional Approach. 2nd Edition. Philadelphia: Evolve Elsevier; 2018.
10. Mathieson S. Does Cambridge University Hospital's Epic project indicate NHS lacks capacity?. 2015. URL: http://www.computerweekly.com/feature/Does-Cambridge-Universitys-Epic-project-indicate-NHS-lacks-capacity [accessed 2018-01-31] [WebCite Cache ID 6wstihTuz]
11. Moore-Colyer R. HP overhauls Cambridge University Hospitals? IT in £200m project. 2015. URL: https://www.v3.co.uk/v3-uk/news/2414532/hp-overhauls-cambridge-university-hospitals-it-in-gbp200m-project [accessed 2018-01-31] [WebCite Cache ID 6wstvlsSZ]
12. Zhang Y, Qiu M, Tsai C, Hassan MM, Alamri A. Health-CPS: healthcare cyber-physical system assisted by cloud and big data. IEEE Systems Journal 2017 Mar;11(1):88-95. [doi: 10.1109/Jsyst.2015.2460747]
13. Peddi SVB, Kuhad P, Yassine A, Pouladzadeh P, Shirmohammadi S, Shirehjini AAN. An intelligent cloud-based data processing broker for mobile e-health multimedia applications. Future Gen Comput Sys 2017 Jan;66:71-86. [doi: 10.1016/j.future.2016.03.019]

14. Tawalbeh L, Bakheder W, Song H. A mobile cloud computing model using the Cloudlet scheme for big data applications. 2016 Presented at: 2016 IEEE First International Conference on Connected Health Applications; Washington.
15. Bussery J, Denis L, Guillon B, Liu P, Marchetti G, Rahal G. eTRIKS platform: conception and operation of a highly scalable cloud-based platform for translational research and applications development. Comput Biol Med 2018 Apr 01;95:99-106. [doi: 10.1016/j.compbiomed.2018.02.006] [Medline: 29482119]
16. Hostetter J, Khanna N, Mandell JC. Integration of a zero-footprint cloud-based picture archiving and communication system with customizable forms for radiology research and education. Acad Radiol 2018 Jun;25(6):811-818. [doi: 10.1016/j.acra.2018.01.031] [Medline: 29555567]
17. Pinheiro A, Dias Canedo E, de Sousa Junior RT, de Oliveira Albuquerque R, García Villalba LJ, Kim T. Security architecture and protocol for trust verifications regarding the integrity of files stored in cloud services. Sensors (Basel) 2018 Mar 02;18(3):753 [FREE Full text] [doi: 10.3390/s18030753] [Medline: 29498641]
18. Cheng Q, Zhang X, Ma J. ICASME: an improved cloud-based authentication scheme for medical environment. J Med Syst 2017 Mar;41(3):44. [doi: 10.1007/s10916-017-0693-8] [Medline: 28181095]
19. Datta S, Bettinger K, Snyder M. Corrigendum: secure cloud computing for genomic data. Nat Biotechnol 2016 Oct 11;34(10):1072. [doi: 10.1038/nbt1016-1072c] [Medline: 27727225]
20. Vidal D, Iriso S, Mulero R. CLARUS as a cloud security framework: eHealth use case. Stud Health Technol Inform 2017;245:1240. [Medline: 29295327]
21. Sahi A, Lai D, Li Y. Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan. Comput Biol Med 2016 Dec 01;78:1-8. [doi: 10.1016/j.compbiomed.2016.09.003] [Medline: 27639895]
22. Glasberg R, Hartmann M, Draheim M, Tamm G, Hessel F. Risks and crises for healthcare providers: the impact of cloud computing. Sci World J 2014;2014:524659 [FREE Full text] [doi: 10.1155/2014/524659] [Medline: 24707207]
23. AbuKhousa E, Mohamed N, Al-Jaroodi J. eHealth cloud: opportunities and challenges. Future Internet 2012 Jul 04;4(4):621-645. [doi: 10.3390/fi4030621]
24. Chiasson M, Davidson E. Pushing the contextual envelope: developing and diffusing IS theory for health information systems research. Information and Organization 2004;14(3):155-188. [doi: 10.1016/j.infoandorg.2004.02]
25. Griebel L, Prokosch H, Köpcke F, Toddenroth D, Christoph J, Leb I, et al. A scoping review of cloud computing in healthcare. BMC Med Inform Decis Mak 2015;15(17):8219 [FREE Full text] [doi: 10.1186/s12911-015-0145-7] [Medline: 25888747]
26. Wand Y, Monarchi DE, Parsons J, Woo CC. Theoretical foundations for conceptual modelling in information systems development. Dec Sup Sys 1995 Dec;15(4):285-304. [doi: 10.1016/0167-9236(94)00043-6]
27. Kebede MM, Liedtke TP, Möllers T, Pischke CR. Characterizing active ingredients of eHealth interventions targeting persons with poorly controlled type 2 diabetes mellitus using the behavior change techniques taxonomy: scoping review. J Med Internet Res 2017 Oct 12;19(10):e348 [FREE Full text] [doi: 10.2196/jmir.7135] [Medline: 29025693]
28. Albahri OS, Albahri AS, Mohammed KI, Zaidan AA, Zaidan BB, Hashim M, et al. Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: taxonomy, open challenges, motivation and recommendations. J Med Syst 2018 Mar 22;42(5):80. [doi: 10.1007/s10916-018-0943-4] [Medline: 29564649]
29. Herzlinger RE. Why innovation in health care is so hard. Harv Bus Rev 2006 May;84(5):58-66. [Medline: 16649698]
30. Rogers E. Diffusion of Innovations. 5th Edition. New York: Free Press; 2003.
31. Linthicum D. Everyone has heard of the cloud, but few know what it is. 2013. URL: http://www.infoworld.com/article/2613402/cloud-computing/everyone-has-heard-of-the-cloud--but-few-know-what-it-is.html [accessed 2018-06-27] [WebCite Cache ID 6wsuv1OEp]
32. Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. Eur J Inf Syst 2012 Jun 19;22(3):336-359. [doi: 10.1057/ejis.2012.26]
33. Moher D. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264. [doi: 10.7326/0003-4819-151-4-200908180-00135]
34. Patton M. Qualitative Research & Evaluation Methods: Integrating Theory and Practice: the Definitive Text of Qualitative Inquiry Frameworks and Options. 4th Edition. Thousand Oaks: Sage Publlications; 2015.
35. Worldwide public cloud services spending forecast to reach $122.5 billion in 2017, according to IDC.: IDC; 2017. URL: https://www.idc.com/getdoc.jsp?containerId=prUS42321417 [accessed 2018-04-16] [WebCite Cache ID 6yjHfKR1N]
36. Kelleher J. Cloud computing market set to increase 21.9 percent annually in the healthcare industry. 2017. URL: https://www.opengovasia.com/articles/6631-cloud-computing-market-set-to-increase-219-percent-annually-in-the-healthcare-industry [accessed 2018-04-17] [WebCite Cache ID 6yk2hnITb]
37. Kshetri N. IT in the Chinese healthcare industry. IT Prof 2013 Jan;15(1):12-15. [doi: 10.1109/MITP.2013.14]
38. Hors-Fraile S, Rivera-Romero O, Schneider F, Fernandez-Luque L, Luna-Perejon F, Civit-Balcells A, et al. Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: a scoping review. Int J Med Inform 2017 Dec 28:1. [doi: 10.1016/j.ijmedinf.2017.12.018] [Medline: 29331276]
39. Schneider JA, Holland CP. eHealth search patterns: a comparison of private and public health care markets using online panel data. J Med Internet Res 2017 Apr 13;19(4):e117 [FREE Full text] [doi: 10.2196/jmir.6739] [Medline: 28408362]

40. Columbus L. Roundup of cloud computing forecasts and market estimates. 2016. URL: http://www.forbes.com/sites/louiscolumbus/2016/03/ [accessed 2018-06-27] [WebCite Cache ID 6wsvbadSy]
41. Yao Q, Han X, Ma X, Xue Y, Chen Y, Li J. Cloud-based hospital information system as a service for grassroots healthcare institutions. J Med Syst 2014 Sep;38(9):104. [doi: 10.1007/s10916-014-0104-3] [Medline: 25015761]
42. Sahni N, Huckman R, Chigurupati A, Cutler D. The IT transformation health care needs. Harvard Bus Rev 2017(Nov–Dec 2017):128-138.
43. Nepal S, Pathan M. Security, Privacy and Trust in Cloud Systems. Berlin: Springer; 2017.
44. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2014 Aug 21;22:E28-E33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
45. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
46. Singh A, Chatterjee K. Cloud security issues and challenges: a survey. J Network Comput App 2017 Feb;79:88-115. [doi: 10.1016/j.jnca.2016.11.027]
47. Moura J, Hutchison D. Review and analysis of networking challenges in cloud computing. J Network Comput App 2016 Jan;60:113-129. [doi: 10.1016/j.jnca.2015.11.015]
48. Attaran M. Cloud computing technology: leveraging the power of the internet to improve business performance. J Int Technol Inform Manag 2017;26(1):6 [FREE Full text]
49. Hassan MM, Lin K, Yue X, Wan J. A multimedia healthcare data sharing approach through cloud-based body area network. Future Gen Comput Sys 2017 Jan;66:48-58. [doi: 10.1016/j.future.2015.12.016]
50. Zhang NJ, Seblega B, Wan T, Unruh L, Agiro A, Miao L. Health information technology adoption in U.S. acute care hospitals. J Med Syst 2013 Apr;37(2):9907. [doi: 10.1007/s10916-012-9907-2] [Medline: 23340826]
51. Gao F, Thiebes S, Sunyaev A. Exploring cloudy collaboration in healthcare: an evaluation framework of cloud computing services for hospitals. 2016 Presented at: 49th Hawaii International Conference on System Sciences (HICSS); 2016; Koloa.
52. Langmead B, Nellore A. Cloud computing for genomic data analysis and collaboration. Nat Rev Genet 2018 Apr;19(4):208-219. [doi: 10.1038/nrg.2017.113] [Medline: 29379135]
53. Grudniewicz A, Bhattacharyya O, McKibbon KA, Straus SE. User-centered design and printed educational materials: a focus group study of primary care physician preferences. J Contin Educ Health Prof 2016;36(4):249-255. [doi: 10.1097/CEH.0000000000000112] [Medline: 28350305]
54. Konduri N, Bastos LGV, Sawyer K, Reciolino LFA. User experience analysis of an eHealth system for tuberculosis in resource-constrained settings: a nine-country comparison. Int J Med Inform 2017 Dec;102:118-129 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.03.017] [Medline: 28495339]
55. Krist AH, Woolf SH. A vision for patient-centered health information systems. JAMA 2011 Jan 19;305(3):300-301. [doi: 10.1001/jama.2010.2011] [Medline: 21245186]
56. Wilson E. Patient-Centered e-Health. Hershey: Medical Information Science Reference; 2009.
57. Scott KC, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
58. Tawalbeh LA, Mehmood R, Benkhlifa E, Song H. Mobile cloud computing model and big data analysis for healthcare applications. IEEE Access 2016;4:6171-6180. [doi: 10.1109/ACCESS.2016.2613278]
59. Åhlfeldt R, Persson A, Rexhepi H, Wåhlander K. Supporting active patient and health care collaboration: a prototype for future health care information systems. Health Informatics J 2016 Dec;22(4):839-853. [doi: 10.1177/1460458215590862] [Medline: 26261220]
60. Rudin RS, Bates DW. Let the left hand know what the right is doing: a vision for care coordination and electronic health records. J Am Med Inform Assoc 2014;21(1):13-16 [FREE Full text] [doi: 10.1136/amiajnl-2013-001737] [Medline: 23785099]
61. Labes S, Hanner N, Zarnekow R. Successful business model types of cloud providers. Bus Inf Syst Eng 2016 Nov 9;59(4):223-233. [doi: 10.1007/s12599-016-0455-z]
62. Benlian A, Kettinger WJ, Sunyaev A, Winkler TJ. The transformative value of cloud computing: a decoupling, platformization, and recombination theoretical framework. J Manag Inform Sys 2018;35(3):1 [FREE Full text]
63. Schneider S, Sunyaev A. Determinant factors of cloud-sourcing decisions: reflecting on the IT outsourcing literature in the era of cloud computing. J Inf Technol 2016;31(1):1-31. [doi: 10.1057/jit.2014.25]
64. Alkhater N, Walters R, Wills G. An empirical study of factors influencing cloud adoption among private sector organisations. Telematics and Informatics 2018 Apr;35(1):38-54. [doi: 10.1016/j.tele.2017.09.017]
65. Eltayeb M, Dawson M. Understanding user's acceptance of personal cloud computing: using the technology acceptance model. In: Latifi S, editor. Information Technology: New Generations. Volume 448. Berlin: Springer; 2016:3-12.
66. Ashtari S, Eydgahi A. Student perceptions of cloud applications effectiveness in higher education. J Comput Sci 2017 Nov;23:173-180. [doi: 10.1016/j.jocs.2016.12.007]

67. Lins S, Schneider S, Sunyaev A. Trust is good, control is better: creating secure clouds by continuous auditing. IEEE Trans Cloud Comput 2016;early access:1. [doi: 10.1109/TCC.2016.2522411]
68. Schneider S, Lansing J, Gao F, Sunyaev A. Development of a taxonomy for cloud service certification criteria: a taxonomic perspective on certification schemes. 2014 Presented at: 47th Hawaii International Conference on System Sciences (HICSS); 2014; Waikoloa p. 6-9.
69. Sunyaev A, Schneider S. Cloud services certification. Commun ACM 2013 Feb 01;56(2):33-36. [doi: 10.1145/2408776.2408789]
70. International standard industrial classification of all economic activities.: United Nations; 2008. URL: https://unstats.un.org/unsd/publication/seriesM/seriesm_4rev4e.pdf [accessed 2018-01-31] [WebCite Cache ID 6wswYAvYp]
71. Poon EG, Jha AK, Christino M, Honour MM, Fernandopulle R, Middleton B, et al. Assessing the level of healthcare information technology adoption in the United States: a snapshot. BMC Med Inform Decis Mak 2006 Jan 05;6:1 [FREE Full text] [doi: 10.1186/1472-6947-6-1] [Medline: 16396679]

Abbreviations

CC: cloud computing
CCS: cloud computing service
IaaS: infrastructure as a service
IT: information technology
PaaS: platform as a service
RQ: research question
SaaS: software as a service

Edited by G Eysenbach; submitted 05.02.18; peer-reviewed by M Lavin, S Albakri, D Gunasekeran; comments to author 11.04.18; accepted 24.04.18; published 11.07.18.

Please cite as:
Gao F, Thiebes S, Sunyaev A
Rethinking the Meaning of Cloud Computing for Health Care: A Taxonomic Perspective and Future Research Directions
J Med Internet Res 2018;20(7):e10041
URL: http://www.jmir.org/2018/7/e10041/
doi:10.2196/10041
PMID:29997108

Original Paper

Device-Embedded Cameras for Eye Tracking–Based Cognitive Assessment: Validation With Paper-Pencil and Computerized Cognitive Composites

Nicholas Bott[1,2], PhD; Erica N Madero[2], MPH; Jordan Glenn[2], PhD; Alexander Lange[2], MSc; John Anderson[2], BS; Doug Newton[2], BA; Adam Brennan[2], BA; Elizabeth A Buffalo[2,3], PhD; Dorene Rentz[4,5], PhD; Stuart Zola[2,6], PhD

[1]Clinical Excellence Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

[2]Neurotrack Technologies, Inc, Redwood City, CA, United States

[3]Department of Physiology and Biophysics, University of Washington, Seattle, WA, United States

[4]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States

[5]Department of Neurology, Brigham and Women's Hospital, Boston, MA, United States

[6]Office of the Provost, Emory University, Atlanta, GA, United States

Corresponding Author:
Nicholas Bott, PhD
Clinical Excellence Research Center
Department of Medicine
Stanford University School of Medicine
75 Alta Road
Stanford, CA, 94306
United States
Phone: 1 650 814 9383
Email: nbott@stanford.edu

Abstract

Background: As eye tracking-based assessment of cognition becomes more widely used in older adults, particularly those at risk for dementia, reliable and scalable methods to collect high-quality data are required. Eye tracking-based cognitive tests that utilize device-embedded cameras have the potential to reach large numbers of people as a screening tool for preclinical cognitive decline. However, to fully validate this approach, more empirical evidence about the comparability of eyetracking-based paradigms to existing cognitive batteries is needed.

Objective: Using a population of clinically normal older adults, we examined the relationship between a 30-minute Visual Paired Comparison (VPC) recognition memory task and cognitive composite indices sensitive to a subtle decline in domains associated with Alzheimer disease. Additionally, the scoring accuracy between software used with a commercial grade eye tracking camera at 60 frames per second (FPS) and a manually scored procedure used with a laptop-embedded web camera (3 FPS) on the VPC task was compared, as well as the relationship between VPC task performance and domain-specific cognitive function.

Methods: A group of 49 clinically normal older adults completed a 30-min VPC recognition memory task with simultaneous recording of eye movements by a commercial-grade eye-tracking camera and a laptop-embedded camera. Relationships between webcam VPC performance and the Preclinical Alzheimer Cognitive Composite (PACC) and National Institutes of Health Toolbox Cognitive Battery (NIHTB-CB) were examined. Inter-rater reliability for manually scored tests was analyzed using Krippendorff's kappa formula, and we used Spearman's Rho correlations to investigate the relationship between VPC performance scores with both cameras. We also examined the relationship between VPC performance with the device-embedded camera and domain-specific cognitive performance.

Results: Modest relationships were seen between mean VPC novelty preference and the PACC (r=.39, *P*=.007) and NIHTB-CB (r=.35, *P*=.03) composite scores, and additional individual neurocognitive task scores including letter fluency (r=.33, *P*=.02), category fluency (r=.36, *P*=.01), and Trail Making Test A (–.40, *P*=.006). Robust relationships were observed between the 60 FPS eye tracker and 3 FPS webcam on both trial-level VPC novelty preference (r=.82, *P*<.001) and overall mean VPC novelty preference (r=.92 *P*<.001). Inter-rater agreement of manually scored web camera data was high (kappa=.84).

Conclusions: In a sample of clinically normal older adults, performance on a 30-minute VPC task correlated modestly with computerized and paper-pencil based cognitive composites that serve as preclinical Alzheimer disease cognitive indices. The strength of these relationships did not differ between camera devices. We suggest that using a device-embedded camera is a reliable and valid way to assess performance on VPC tasks accurately and that these tasks correlate with existing cognitive composites.

(J Med Internet Res 2018;20(7):e11143) doi:10.2196/11143

KEYWORDS

eye tracking; visual paired comparison; preclinical Alzheimer's disease; neuropsychological testing

Introduction

Alzheimer disease (AD) and other forms of dementia, broadly characterized by declines in mental ability severe enough to interfere with daily life, pose serious challenges to patients, caregivers, and healthcare systems worldwide. As populations age, the global prevalence of dementia is expected to triple to 132 million between 2015 and 2050 [1]. In the United States (US) alone, the costs of AD are projected to grow by 400% from US $186 billion in 2018 to US $750 billion in 2050 as the number of people with dementia increases from 5.5 million to 13.8 million [2]. Alzheimer disease can go undetected for long periods of time because the disease has a prolonged preclinical phase, during which neuronal and neurobiological changes can occur for years or decades before noticeable symptoms appear. Early detection of AD during the preclinical phase has the potential to decrease medical and long-term care costs by as much as US $7 trillion in the US [2]. Detection of preclinical AD can enable people to seek treatment earlier, address modifiable risk factors, and potentially slow the progression of the disease, ultimately preserving cognitive function and reducing population health care costs [1,2].

Current detection methods for preclinical AD include the use of biomarkers, such as neuroimaging and cerebrospinal fluid tests for amyloid-β and tau proteins [3,4]. Increasingly, cognitive assessment composites targeting relevant cognitive domains sensitive to AD pathologies, such as the Preclinical Alzheimer's Cognitive Composite (PACC) [5] and National Institutes of Health Toolbox Cognitive Battery (NIHTB-CB) [6,7], have shown efficacy in stratifying preclinical AD populations. However, there are drawbacks to both detection methods that ultimately limit their feasibility for screening large populations. For example, biomarker tests are expensive and invasive, cognitive batteries require trained staff for standardized administration, and both methods restrict access by requiring people to travel to a clinic.

Another method for detecting presymptomatic cognitive decline is through the use of eye tracking systems that assess eye movement behavior [8]. For example, eye tracking–based tasks that assess saccade patterns can be used to detect mild cognitive impairment (MCI) and AD [9,10]. Saccades, or small rapid movement of the eyes between fixations of relevant stimuli, can be examined within certain task paradigms to quantify inhibitory control. Another eye tracking-based task of particular interest is the visual paired-comparison (VPC) task, which has been shown to reliably detect early signs of cognitive decline in older adults before symptoms are present [11,12]. Visual paired-comparison tasks are a well-established method for detecting memory dysfunction in humans and other primates, from infancy through adulthood [11-18].

Commercial-grade eye-tracking cameras have traditionally been used to collect VPC data. These high frame rate cameras can capture a variety of complex visual features, including saccades, smooth pursuit (consistent tracking movements of the eyes), and fixation distributions (eyes focusing on particular areas or items), which can all be assessed for abnormal patterns indicative of a variety of pathologies [19,20]. Data can either be analyzed automatically with software provided by the manufacturer or inspected manually by researchers who have experience evaluating eye-tracking metrics. However, the use of commercial eye trackers as a cognitive health screening tool has limitations, much like biomarker tests and traditional cognitive assessments. Eye-tracking devices are expensive, complicated to use, and are not widely available in clinical settings. To effectively reach the growing number of individuals at risk for cognitive decline, preclinical AD screening tests that are both reliable and scalable, such as VPC tests that utilize ubiquitous device-embedded webcams (eg, mobile phones, tablets, laptops), need to be validated and implemented.

Emerging research indicates that when used for cognitive tests sensitive to early signs of decline, embedded webcams in laptops and mobile devices produce data of similar quality to that collected by commercial-grade cameras [21]. It remains to be determined if scoring accuracy will be maintained for a longer (30 minutes) version of the test. The 30-minute VPC task has been shown to reliably predict future declines in cognitive status among clinically normal individuals and individuals with amnestic MCI, a subtype of MCI with focal deficits in learning and memory performance [12]. The validation of accurate scoring methods for the embedded camera version of the test and its relationship to paper-pencil and computerized cognitive composites would add considerable value to the task as an asset to a clinician's assessment repertoire.

The purpose of this study is (1) to investigate the relationships between performance on a 30-minute webcam-based digital VPC task and two cognitive composite indices sensitive to subtle impairment in AD-relevant cognitive domains, (2) to examine the relationship between performance on the VPC task with a device-embedded camera and domain-specific cognitive scores, and (3) to investigate the accuracy of human-coded gaze positions on a thirty-minute VPC using a laptop-embedded camera when compared to an automatically scored gold standard high frame rate eye-tracking camera.

Methods

Participants and Procedures

All subjects underwent informed consent procedures approved by the Partners Human Research Committee, the Institutional Review Board for Brigham and Women's Hospital and Massachusetts General Hospital. A total of 49 clinically normal, community-dwelling older adults were recruited from a cohort of volunteers interested in participating in research studies at the Center for Alzheimer Research and Treatment at Brigham and Women's Hospital and the Massachusetts Alzheimer Disease Research Center at Massachusetts General Hospital. Subjects were excluded if they had a history of alcoholism, drug abuse, head trauma, or current serious medical or psychiatric illnesses. All subjects above the age of 50 years and within age-specified norms on the Telephone Interview of Cognitive Status [22] were eligible for the study. No prior computer or iPad knowledge was required to participate. Subjects attended 1 clinic visit, during which they completed paper-pencil based cognitive tasks including the PACC, the NIHTB-CB, and the Neurotrack 30-minute VPC eye-tracking assessment. Eye-tracking data for the 30-minute VPC task was collected simultaneously by a commercial-grade eye tracker and a laptop-embedded camera.

Cognitive Composites

The PACC is a paper-pencil cognitive composite that includes 2 tasks of episodic memory, a task of speeded executive functioning, and a global cognitive screen. The Logical Memory-delayed recall score and the Free and Cued Selective Reminding Test total score comprised the episodic memory tests, with the Wechsler Adult Intelligence Scale-Revised Digit Symbol Coding Test total score representing the speeded executive functioning measure, and the Mini-Mental State Exam total score serving as the global cognitive screen [23].

Additionally, a measure of attention and processing speed (Trail Making Test A) and 2 measures of executive functioning (letter and verbal fluency) were administered. All tests were z-transformed using the performance means and standard deviations of clinically normal older adults (n=256, age range: 61-90 years) [24,25]. All four z-transformed variables were averaged together to produce a PACC composite score, with higher scores indicating better performance.

The NIHTB-CB is a computerized cognitive composite comprised of the Picture Vocabulary Test (PVT), the Flanker Inhibitory Control and Attention Test (Flanker), the Dimensional Change Card Sort Test (DCCS), the Pattern Comparison Processing Speed Test (PCPST), and the Picture Sequence Memory Test (PSMT) [7]. The PVT is a measure of receptive vocabulary, requiring participants to select from 4 images the 1 closest to the meaning of an orally presented word. The Flanker is a measure of cognitive control, requiring participants to focus on a stimulus surrounded by 4 identical stimuli around the target and having them select the direction in which the target stimulus is pointing. The DCCS is a measure of executive control, requiring participants to shift set matching a target visual stimulus to stimuli by shape or color. The PCPST is a measure of processing speed, requiring participants to rapidly match an object by shape or color. The PSMT is a measure of visual episodic memory, requiring participants to re-create the order of a set of images over 2 test trials [7]. The Flanker, PCPST, DCCS, and PSMT were scored per NIHTB-CB guidelines, and overall performance was quantified by a theta score, calculated by combining all of the scores on the individual tasks.

Visual Paired-Comparison Test Construction

A 30-minute VPC task developed by Neurotrack Technologies Inc (Redwood City, CA) was used in this study. VPC tasks quantify how the test subject splits attention between familiar and novel visual stimuli, with a familiarization phase preceding a testing phase. During the familiarization phase, subjects were presented with pairs of identical visual stimuli for a fixed period (5 seconds). During the test phase, which follows a delay of either 2 seconds or 2 minutes to assess immediate and delayed recognition memory, subjects were presented with additional pairs of visual stimuli, including 1 from the familiarization phase (familiar stimulus) and 1 novel stimulus. The ratio of time a subject spends gazing at the novel stimulus relative to the total viewing time produces a novelty preference (NP) score, with higher scores representing better declarative memory function and lower scores indicating impaired function [11,26,27].

System Components

Eye movements during the VPC task were simultaneously recorded with a commercial-grade Tobii X2-60 eye tracker camera system (Tobii AB, Stockholm, Sweden) and an embedded web camera on a 13-inch Apple MacBook Air laptop (Apple, Cupertino, CA). The Tobii camera sampled at 60 Hz, with corneal and pupil centers determining the gaze angle. Eye data were recorded using the Tobii SDK and API software. Participants were seated approximately 27 inches from the 13-inch laptop monitor that displayed the visual stimuli. The Apple MacBook Air laptop processor was a 1.4 GHz Intel Core i5 with 4 GB 1,600 MHz DDR3 memory and a 1,536 MB Intel HD Graphics 5,000 Graphics card. Video resolution of the laptop during test recording was 640 by 480.

Calibration Validation and Gaze Position

Explanations regarding the validation of camera calibration, data acquisition, and fixation filters for device-embedded cameras have previously been reported [21]. Briefly, before the start of the VPC task, subjects were instructed to watch a blue dot travel around the screen. Acting as a coordinate system, the top left of the screen represented (0, 0) and the bottom right of the screen represented (1, 1). The calibration ball traveled a predetermined path: (0.5, 0.5), (0.1, 0.1), (0.1, 0.9), (0.9, 0.9), (0.5, 0.5), (0.9, 0.1), (0.5, 0.1), (0.1, 0.5), (0.5, 0.9), pausing at each of the above points for approximately 2 seconds. Calibration validation of the device-embedded camera was determined by three human coders evaluating the individual frames of the calibration-phase video. Coding of the calibration phase video was repeated if individual accuracy of correctly coded calibration frames was below 90%. Calibration data were used to generate individualized models to predict gaze location and duration but were not incorporated into the experimental

procedure. Calibration validation of the commercial grade eye-tracking camera was determined by multiple accuracy metrics produced by the Tobii X2-60 SDK/API software.

Using the Tobii Pro Analytics software development kit default Tobii Pro Studio settings, we utilized the Active Display Coordinate System and the User Coordinate System to determine gaze location. Each data point consisted of an estimated gaze point for both the left and right eye. For each data point, the midpoint of the 2 gaze points was used as the definitive gaze estimate. Test trials were automatically excluded if more than 4s of data was missing due to the Tobii failing to find the eyes of the subject. In a previous study, Zola and colleagues used an Applied Science Laboratories eye tracker that recorded gaze data at 120 Hz (120 frames per second) [12]. To replicate the cluster-based algorithm used by Zola et al [12], a fixation filter to process the raw Tobii data was developed [21]. Three researchers with expertise in eye-tracking behavior and the Tobii X2-60 eye tracker system independently inspected all test trials to ensure the quality of test data. Test trials flagged for aberrant gaze paths (eg, gaze clustering, erratic saccades) were discussed corporately, and a consensus decision was made to retain or discard the trial in question.

In addition to the commercial eye tracker, subjects were simultaneously recorded with a device embedded camera during the calibration and test phases. A high definition Flash video recorder recorded the subject, and the resulting Flash video (FLV) footage was streamed to Neurotrack's Wowza Amazon Web Services instance. Metadata, such as calibration phase timing and timing of task image presentation, was injected into the FLV video to ensure correspondence between frames of the video and events of the test.

Scoring

Performance on preferential looking VPC tasks is quantified as novelty preference. In the present study, novelty preference was defined as the percentage of time the participant spent looking at the novel image compared with the familiar image. For each test trial, NP was calculated as (time viewing novel image) divided by (total time viewing either image). Mean novelty preference for each of the 20 test trials yielded the overall novelty preference score. Using the commercial grade eye-tracking camera, a rectangular area of interest perimeter slightly larger than each image was defined. Gaze time on each image was calculated based on the total gaze fixation time recorded by the Tobii X2-60 software.

For the device-embedded camera data, individual processed video frames were evaluated on a frame by frame basis down-sampled to 3 frames per second (FPS) by 3 independent human coders to determine whether the subject was looking to the left, right, or neither side of the screen. Coding of the "neither" option was intended for frames when the participant was blinking, or when the image was of poor enough quality that the iris was indistinguishable from the rest of the eye. For each image, the majority decision was taken by the individual ratings. The NP score for each trial was the percentage of frames that the participant was rated as looking at the novel side (no. of "novel" frames) divided by (total no. of "novel" frames + no. of "familiar" frames).

Visual Paired-Comparison Data Analysis

Analyses of VPC test data were conducted with IBM SPSS version 24.0 using non-parametric statistical procedures due to the non-normal distribution of VPC test performance. Inter-rater agreement of web camera data scoring was assessed using Krippendorff's kappa calculation [28]. Relationships between the Tobii X2-60 eye-tracking camera (60 FPS) and the laptop embedded camera (3 FPS) were assessed using Spearman's Rho correlations. Relationships between VPC task performance, paper-pencil based neuropsychological tasks and computerized neuropsychological tasks were assessed using two-tailed Spearman's Rho correlations. There were no relationships between age, gender or education on the VPC task or individual paper-pencil based neuropsychological tasks. Performance on computerized neuropsychological tasks on the NIHTB-CB was assessed based on standardized scores. The Cohen standard was used to determine the strength of these relationships with correlation coefficients of .10 as weak, .30 as moderate, and .50 and above as strong [29]. The strength of inter-rater reliability kappa statistic was determined with reliability of .40 to .59 as weak, .60 to .79 as moderate, and .80 to .90 as strong.

Results

Participant Characteristics

Subjects were all cognitively normal community-dwelling older adults. The age range of the study cohort was 54-97 years and the level of education ranged from 12-20 years. (Table 1).

Table 1. Characteristics of the study cohort of cognitively normal older adults.

Characteristic	Value (N=49)
Age (years), mean (SD)	69 (8)
Gender, n (%)	
Female	28 (57)
Male	21 (43)
Years of education, mean (SD)	16 (3)
Race, n (%)	
European-American	31 (63)
African American	18 (37)

Table 2. Correlations between visual paired comparison task performance by camera type and cognitive assessments.

Cognitive assessment	Tobii X2-60	*P* value[a]	Laptop-embedded Camera	*P* value[a]	Fisher r to z transformation (*P* value)
PACC[d] Composite	.43	.005[b]	.39	.007[b]	.42
MMSE[e]	.20	.21	.13	.40	.37
LM-DR[f]	.30	.06	.25	.09	.16
FCSRT[g]	–.05	.75	.22	.15	.11
Digit Symbol Coding	.48	.001[b]	.32	.03[c]	.18
Letter Fluency	.44	.004[b]	.33	.02[c]	.28
Category Fluency	.43	.005[b]	.36	.01[c]	.36
Trails A	–.45	.003[b]	–.40	.006[b]	.39
NIHTB-CB[h] Composite	.32	.049[c]	.35	.03[c]	.40
PVT[i]	.28	.09	.28	.07	.50
PSMT[j]	.39	.01[c]	.33	.03[c]	.38
Flanker	–.006	.97	.22	.16	.16
DCCS[k]	.29	.07	.37	.02[c]	.35

[a]Determined by Spearman correlations and two-tailed Fisher r to z correlation comparisons.

[b]*P*<.01.

[c]*P*<.05.

[d]PACC: : Preclinical Alzheimer's Cognitive Composite.

[e]MMSE: Mini-Mental State Exam.

[f]LM-DR: Logical Memory-delayed recall.

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

[h]NIHTB-CB: National Institute of Health Toolbox Cognitive Battery.

[i]PVT: Picture Vocabulary Test.

[j]PSMT: Picture Sequence Memory Test.

[k]DCCS: Dimensional Change Card Sort.

Correlations Between Visual Paired-Comparison Data and Cognitive Composites

To further investigate the data from the commercial-grade eye-tracking camera and the laptop-embedded camera, we compared the strength of the correlations between the Tobii X2-60 VPC and the PACC and NIHTB-CB composites with the strength of the correlations between the laptop embedded camera and the PACC and NIHTB-CB composites. Fisher r to z transformation revealed no significant differences in the strength of correlation between each modality of data acquisition (*P*>.10; Table 2).

Associations Between Visual Paired-Comparison Data and Cognitive Domains

Performance on the NIHTB-CB was moderately correlated with scores on the PACC (r=.51, *P*<.001), which is in line with what has been published previously [6]. We also examined the correlation between VPC task performance with device-embedded camera data and cognitive test batteries. Analyses found modest relationships between VPC task performance and the PACC (r=.39, *P*=.007) and the NIHTB-CB (r=.35, *P*=.03) across 46 subjects. Three subjects were excluded from the analysis due to insufficient data quality.

We then analyzed the correlations between VPC performance and domain-specific cognitive functions (Table 2). Significant relationships were observed on digit symbol coding (r=.32, *P*=.03), Trails A (r=–.40, *P*=.006), letter fluency (r=0.33, *P*=.02), and category fluency (r=0.36, *P*=.01). A trend relationship was seen on Logical Memory Delayed Recall (r=.25, *P*=.09). On the NIHTB-CB, significant relationships were observed on PSMT (r=.33, *P*=.03) and DCCS (r=.37, *P*=.02), with a trend relationship for PVT (r=.28, *P*=.07).

Visual Paired-Comparison Data Scoring Correlations

Analysis of the relationship between data from the commercial grade eye-tracker and the device-embedded web camera revealed strong positive associations overall. Spearman's Rho correlation was .91 (n=44, *P*<.001) among study participants (Figure 1).

Figure 1. Relationship between overall mean novelty preference scores collected by the Tobii commercial-grade eye tracker and device-embedded web camera.

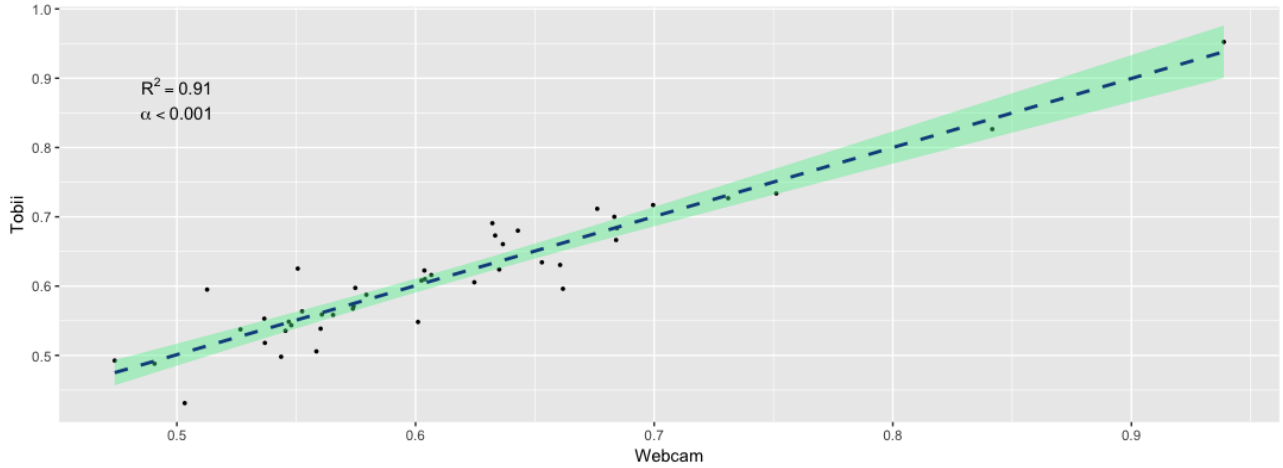
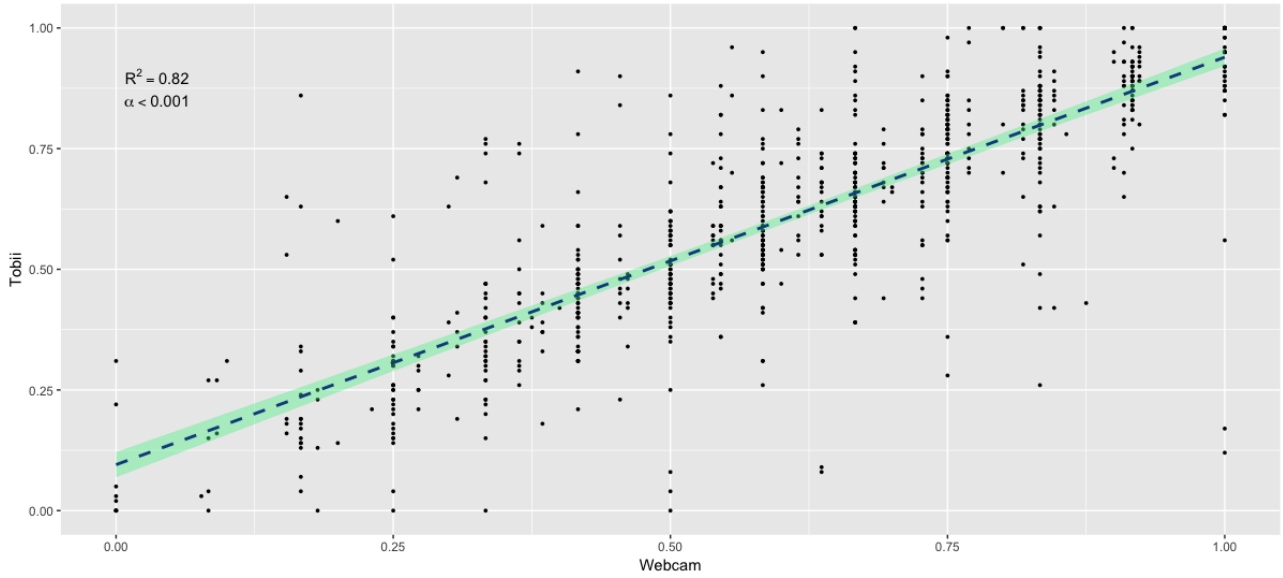


Figure 2. Relationship between trial-level novelty preference scores collected by the Tobii commercial-grade eye tracker and device-embedded web camera.



Next, we compared the relationship between data from the commercial grade eye tracker camera and the laptop-embedded camera for each of the 20 test trials per participant. Analyses revealed robust associations between each camera type across test trials, with a Spearman's Rho correlation of .82 (n=841, *P*<.001; Figure 2). Inter-rater reliability of scoring the laptop embedded camera data using Krippendorff's kappa formula revealed a strong agreement between the three human raters for each of the 15 frames across each of the 20 test trials (kappa=.84).

Discussion

The primary focus of this study was to examine the relationships between VPC performance and traditional cognitive assessments known to be sensitive to signs of early cognitive dysfunction. Previous studies have shown differential performance on the 30-minute VPC task between various cognitive subgroups [11], as well as the predictive value of the task in identifying individuals who will progress from normal cognitive function to amnestic MCI (aMCI) or from aMCI to AD within three years of the assessment [12]. However, these previous studies only collected data with commercial-grade eye trackers and did not investigate the correlation between VPC performance and cognitive composite scores. Our results demonstrate convergent validity between a 30-minute VPC eye-tracking task and both the PACC and NIHTB-CB batteries. This investigation presents the first data demonstrating modest to moderate correlations between VPC task performance using device-embedded cameras and scores on gold standard cognitive composites, suggesting these eye-tracking-based tests can provide complementary support to conventional cognitive composites for detecting early cognitive changes.

We also discovered that the observed correlations between VPC performance and cognitive battery scores were driven by particular cognitive domains. Specifically, VPC performance correlated the highest with measures of processing speed, executive function, and visual episodic memory. While a trend association was seen on a measure of verbal episodic memory, the exclusively visual nature of the VPC task would be expected to drive a stronger relationship with other measures of visual

episodic memory. VPC task performance has previously been shown to be associated with processing speed [30], which accounts for a large proportion of variance across cognitive tasks, including executive functioning [31].

The last major focus of this study was to explore the relationship between VPC performance data collected by distinct camera types. To our knowledge, this is the first study to measure the correlations between VPC task data collected from both commercial-grade and device-embedded cameras for a 30-minute VPC task. Commercial-grade eye-tracking technologies have been shown to detect abnormal eye movements across a number of clinical populations, including people with schizophrenia [32], autism [33], ADHD [34], multiple sclerosis [35], and cognitive decline [36]. These high frame rate cameras typically collect an abundance of data on eye movement behavior, including saccades, gaze fixation and duration, smooth pursuit, and other metrics that can provide valuable insights for certain cognitive processes [19,20]. However, these devices are used primarily in research settings due to their complexity and high cost, so the need exists for an alternative eye-tracking system that is feasible for more widespread use. This study demonstrates that data collected from webcams at 30 FPS that is subsequently down-sampled to 3 FPS can provide clinically relevant insights into cognitive function. As such, the robust datasets collected by commercial-grade (ie, 60 FPS) cameras are not always necessary for certain assessments, such as VPC tasks.

Research on methods for the recording and analyzing eye movements from device-embedded web cameras continues to grow [37-44], demonstrating the utility of real-time online systems and offline recording systems. Perhaps the most significant advantage of built-in web cameras is their lack of geographical restriction to collect eye feature data on large samples sizes. For example, open source eye-tracking software, such as WebGazer.js [45] can be deployed across most major web browsers to provide insight into the eye movements of website visitors. The widespread reach of device-embedded cameras has the potential to greatly increase access to eye-tracking-based cognitive assessments across geographically dispersed populations, as people can take the tests anytime in their own homes.

These results further demonstrate that both commercial-grade eye trackers and device-embedded cameras can produce robust data of sufficient quality for analyzing VPC task performance. High correlations existed between VPC performance using commercial-grade devices and device-embedded cameras at both the overall and trial level, suggesting that webcams represent a consistent, scalable, and reliable method for VPC data collection. These findings align with results from a previous study, in which we demonstrated strong associations between manually scored data from a device-embedded camera and automatically scored data from a commercial-grade eye tracker for an abbreviated 5-minute version of the VPC task [21]. The growing evidence base supporting the comparability of VPC data between commercial-grade and embedded cameras is an important development for the field of remote cognitive assessments.

The scalability and lower cost of the webcam-based VPC task holds the potential to greatly increase screening rates for early signs of cognitive decline, which will be an important component of caring for ever-growing aging populations worldwide in the coming decades. While the gold standard PACC and NIHTB-CB cognitive assessments are reliable for detecting preclinical cognitive decline, they are also limited in their scalability, much like commercial-grade eye-tracking devices. The cognitive composites must be taken in person, to ensure standardized administration by a trained professional who can guide participants through the various sets of instructions. Conversely, the webcam-enabled VPC task is better suited for widespread adoption because it is reliable, language-agnostic, requires little to no instruction and minimal equipment, and can be administered and completed anywhere. The VPC task using eye-tracking data collected from web cameras is a potential complement to traditional test batteries for cognitive decline. The further integration and development of these scalable tasks by companies like Neurotrack Technologies, Inc, (Redwood City, CA) will greatly increase the availability of these assessments.

This study has a few limitations. For one, the small study sample comprised of clinically normal older adults restricts the generalizability of the results to broader populations. However, we were able to recruit a diverse group of participants and will strive to do so in larger studies in the future to maximize the external validity of the results. Also, the collection of the webcam-based VPC data within a clinic setting is not ideal for approximating in-home performance, but the validation of webcam-based data in a research setting is a necessary precursor to remote data collection.

These results set the stage for many future directions. We demonstrated here that manual scoring of the webcam-based VPC task had high inter-rater reliability, indicating that this method of data quantification produces consistent results across different scorers. The future development of an automated scoring system for device-embedded camera data would be extremely valuable, allowing for faster scoring and deployment on a larger scale. Additionally, although outside of the scope of this study, future studies will need to examine the test-retest reliability of the webcam-based VPC tests to ensure high internal validity.

In conclusion, this study showed strong convergence in data accuracy between commercial-grade eye tracking cameras and device-embedded cameras on a 30-min VPC task. Results demonstrated modest to moderate correlations on 30-minute VPC task performance using device-embedded cameras and performance on gold standard digital and paper-pencil cognitive composites. Eye tracking through device-embedded cameras can provide efficient and scalable evaluation of cognitive performance and support the growing number of individuals at risk for cognitive decline.

Conflicts of Interest

NB, ENM, JG, AL, JA, DN, and AB are all employees of Neurotrack Technologies, Inc. DR has served as a paid consultant for Eli Lilly, Janssen Pharmaceuticals and Biogen Idec and also serves on the Scientific Advisory Board for Neurotrack Technologies, Inc. EAB and SZ are cofounders of Neurotrack Technologies, Inc. Neurotrack Technologies, Inc, funded this study.

References

1. Alzheimer's Disease International. 2016. World Alzheimer Report 2016: Improving Healthcare for People Living with Dementia URL: https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf [accessed 2018-05-14] [WebCite Cache ID 6zQ29A3aH]
2. Alzheimer's Association. Alzheimers Dement; 2018. Alzheimer's Disease Facts and Figures URL: https://www.alz.org/documents_custom/2018-facts-and-figures.pdf [WebCite Cache ID 6zQ2BpnCh]
3. Tan C, Yu J, Tan L. Biomarkers for preclinical Alzheimer's disease. J Alzheimers Dis 2014 Oct;42(4):1051-1069. [doi: 10.3233/JAD-140843] [Medline: 25024325]
4. Ho JK, Nation DA, Alzheimer's Disease Neuroimaging Initiative. Neuropsychological Profiles and Trajectories in Preclinical Alzheimer's Disease. J Int Neuropsychol Soc 2018 Apr 30:1-10. [doi: 10.1017/S135561771800022X] [Medline: 29706146]
5. Donohue MC, Sperling RA, Petersen R, Sun C, Weiner MW, Aisen PS, Alzheimer's Disease Neuroimaging Initiative. Association Between Elevated Brain Amyloid and Subsequent Cognitive Decline Among Cognitively Normal Persons. JAMA 2017 Dec 13;317(22):2305-2316 [FREE Full text] [doi: 10.1001/jama.2017.6669] [Medline: 28609533]
6. Buckley RF, Sparks KP, Papp KV, Dekhtyar M, Martin C, Burnham S, et al. Computerized Cognitive Testing for Use in Clinical Trials: A Comparison of the NIH Toolbox and Cogstate C3 Batteries. J Prev Alzheimers Dis 2017;4(1):3-11 [FREE Full text] [doi: 10.14283/jpad.2017.1] [Medline: 29188853]
7. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. Cognition assessment using the NIH Toolbox. Neurology 2013 Mar 12;80(11 Suppl 3):S54-S64 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872ded] [Medline: 23479546]
8. Seligman SC, Giovannetti T. The Potential Utility of Eye Movements in the Detection and Characterization of Everyday Functional Difficulties in Mild Cognitive Impairment. Neuropsychol Rev 2015 Jun;25(2):199-215. [doi: 10.1007/s11065-015-9283-z] [Medline: 25851239]
9. Peltsch A, Hemraj A, Garcia A, Munoz DP. Saccade deficits in amnestic mild cognitive impairment resemble mild Alzheimer's disease. Eur J Neurosci 2014 Jun;39(11):2000-2013. [doi: 10.1111/ejn.12617] [Medline: 24890471]
10. Heuer HW, Mirsky JB, Kong EL, Dickerson BC, Miller BL, Kramer JH, et al. Antisaccade task reflects cortical involvement in mild cognitive impairment. Neurology 2013 Oct 01;81(14):1235-1243 [FREE Full text] [doi: 10.1212/WNL.0b013e3182a6cbfe] [Medline: 23986300]
11. Crutcher MD, Calhoun-Haney R, Manzanares CM, Lah JJ, Levey AI, Zola SM. Eye tracking during a visual paired comparison task as a predictor of early dementia. Am J Alzheimers Dis Other Demen 2009 Feb;24(3):258-266 [FREE Full text] [doi: 10.1177/1533317509332093] [Medline: 19246573]
12. Zola SM, Manzanares CM, Clopton P, Lah JJ, Levey AI. A behavioral task predicts conversion to mild cognitive impairment and Alzheimer's disease. Am J Alzheimers Dis Other Demen 2013 Mar;28(2):179-184 [FREE Full text] [doi: 10.1177/1533317512470484] [Medline: 23271330]
13. Gunderson VM, Sackett GP. Development of pattern recognition in infant pigtailed macaques (Macaca nemestrina). Developmental Psychology 1984;20(3):418-426. [doi: 10.1037/0012-1649.20.3.418]
14. Gunderson VM, Swartz KB. Visual recognition in infant pigtailed macaques after a 24-hour delay. Am. J. Primatol 1985;8(3):259-264. [doi: 10.1002/ajp.1350080309]
15. Bachevalier J, Brickson M, Hagger C. Limbic-dependent recognition memory in monkeys develops early in infancy. NeuroReport 1993;4(1):77-80. [doi: 10.1097/00001756-199301000-00020]
16. Pascalis O, de Haan M, Nelson CA, de Schonen S. Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. Journal of Experimental Psychology: Learning, Memory, and Cognition 1998;24(1):249-260. [doi: 10.1037/0278-7393.24.1.249]
17. Manns JR, Stark CE, Squire LR. The visual paired-comparison task as a measure of declarative memory. Proc Natl Acad Sci U S A 2000 Oct 24;97(22):12375-12379 [FREE Full text] [doi: 10.1073/pnas.220398097] [Medline: 11027310]
18. Clark RE, Zola SM, Squire LR. Impaired recognition memory in rats after damage to the hippocampus. J Neurosci 2000 Dec 01;20(23):8853-8860 [FREE Full text] [Medline: 11102494]
19. Gooding DC, Miller MD, Kwapil TR. Smooth pursuit eye tracking and visual fixation in psychosis-prone individuals. Psychiatry Research 2000 Feb;93(1):41-54. [doi: 10.1016/S0165-1781(00)00113-X]
20. Ettinger U, Kumari V, Crawford TJ, Davis RE, Sharma T, Corr PJ. Reliability of smooth pursuit, fixation, and saccadic eye movements. Psychophysiology 2003 Jul;40(4):620-628. [Medline: 14570169]
21. Bott NT, Lange A, Rentz D, Buffalo E, Clopton P, Zola S. Web Camera Based Eye Tracking to Assess Visual Memory on a Visual Paired Comparison Task. Front Neurosci 2017 Jun;11:370 [FREE Full text] [doi: 10.3389/fnins.2017.00370] [Medline: 28701915]

22. Knopman DS, Roberts RO, Geda YE, Pankratz VS, Christianson TJH, Petersen RC, et al. Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. Neuroepidemiology 2010;34(1):34-42 [FREE Full text] [doi: 10.1159/000255464] [Medline: 19893327]
23. Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, Australian Imaging, Biomarkers,Lifestyle Flagship Study of Ageing, Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Cooperative Study. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol 2014 Aug;71(8):961-970 [FREE Full text] [doi: 10.1001/jamaneurol.2014.803] [Medline: 24886908]
24. Dagley A, LaPoint M, Huijbers W, Hedden T, McLaren DG, Chatwal JP, et al. Harvard Aging Brain Study: Dataset and accessibility. Neuroimage 2017 Jan;144(Pt B):255-258 [FREE Full text] [doi: 10.1016/j.neuroimage.2015.03.069] [Medline: 25843019]
25. Hedden T, Mormino EC, Amariglio RE, Younger AP, Schultz AP, Becker JA, et al. Cognitive profile of amyloid burden and white matter hyperintensities in cognitively normal older adults. J Neurosci 2012 Nov 14;32(46):16233-16242 [FREE Full text] [doi: 10.1523/JNEUROSCI.2462-12.2012] [Medline: 23152607]
26. Fantz RL. Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones. Science 1964 Oct 30;146(3644):668-670. [doi: 10.1126/science.146.3644.668]
27. Fagan JF. Memory in the infant. Journal of Experimental Child Psychology 1970 Apr;9(2):217-226. [doi: 10.1016/0022-0965(70)90087-1]
28. Krippendorff K. Content Analysis An Introduction to Its Methodology. In: Content Analysis An Introduction to Its Methodology. Thousand Oaks, CA: SAGE Publications; 2013:A.
29. Cohen J. A power primer. Psychological Bulletin 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
30. Rose SA. Enhancing visual recognition memory in preterm infants. Developmental Psychology 1980;16(2):85-92. [doi: 10.1037/0012-1649.16.2.85]
31. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev 1996 Jul;103(3):403-428. [Medline: 8759042]
32. Holzman PS, Proctor LR, Hughes DW. Eye-Tracking Patterns in Schizophrenia. Science 1973 Jul 13;181(4095):179-181. [doi: 10.1126/science.181.4095.179]
33. Boraston Z, Blakemore S. The application of eye-tracking technology in the study of autism. J Physiol 2007 Jun 15;581(Pt 3):893-898 [FREE Full text] [doi: 10.1113/jphysiol.2007.133587] [Medline: 17430985]
34. Ross RG, Harris JG, Olincy A, Radant A. Eye movement task measures inhibition and spatial working memory in adults with schizophrenia, ADHD, and a normal comparison group. Psychiatry Research 2000 Jul;95(1):35-42. [doi: 10.1016/S0165-1781(00)00153-0]
35. Serra A, Chisari CG, Matta M. Eye Movement Abnormalities in Multiple Sclerosis: Pathogenesis, Modeling, and Treatment. Front Neurol 2018 Feb;9:31 [FREE Full text] [doi: 10.3389/fneur.2018.00031] [Medline: 29467711]
36. Daffner KR, Scinto LF, Weintraub S, Guinessey JE, Mesulam MM. Diminished curiosity in patients with probable Alzheimer's disease as measured by exploratory eye movements. Neurology 1992 Feb;42(2):320-328. [Medline: 1736159]
37. Wang J, Sung E. Gaze determination via images of irises. Image and Vision Computing 2001 Oct;19(12):891-911. [doi: 10.1016/S0262-8856(01)00051-8]
38. Hansen DW, Pece AE. Eye tracking in the wild. Computer Vision and Image Understanding 2005 Apr;98(1):155-181. [doi: 10.1016/j.cviu.2004.07.013]
39. Vivero V, Barriera N, Penedo M, Cabrero D, Remeseiro B. Directional gaze analysis in webcam video sequences. In: Lecture Notes in Computer Science. Vol 6112. Berlinpringer; 2010:316-324.
40. Anderson NC, Risko EF, Kingstone A. Exploiting human sensitivity to gaze for tracking the eyes. Behav Res Methods 2011 Sep;43(3):843-852. [doi: 10.3758/s13428-011-0078-8] [Medline: 21461633]
41. Lin Y, Lin R, Lin Y, Lee GC. Real-time eye-gaze estimation using a low-resolution webcam. Multimed Tools Appl 2012 Aug 14;65(3):543-568. [doi: 10.1007/s11042-012-1202-1]
42. Petridis S, Giannakopoulos T, Spyropoulos CD. Unobtrusive low cost pupil size measurements using web cameras. NetMedicine 2013:9-20.
43. Adhikari S, Stark D. Video-based eye tracking for neuropsychiatric assessment. Ann N Y Acad Sci 2017 Dec;1387(1):145-152. [doi: 10.1111/nyas.13305] [Medline: 28122120]
44. Semmelmann K, Weigelt S. Online webcam-based eye tracking in cognitive science: A first look. Behav Res Methods 2018 Apr;50(2):451-465. [doi: 10.3758/s13428-017-0913-7] [Medline: 28593605]
45. Democratizing Webcam Eye Tracking on the Browser. ? URL: https://webgazer.cs.brown.edu/ [accessed 2018-05-14] [WebCite Cache ID 6zPzODr8T]

Abbreviations

AD: Alzheimer disease
aMCI: amnestic mild cognitive impairment
DCCS: Dimensional Change Card Sort Test

FLV: flash video
FPS: frames per second
MCI: mild cognitive impairment
NIHTB-CB: National Institutes of Health Toolbox Cognitive Battery
NP: novelty preference
PACC: Preclinical Alzheimer's Cognitive Composite
PCPST: Pattern Comparison Processing Speed Test
PSMT: Picture Sequence Memory Test
PVT: Picture Vocabulary Test
VPC: visual paired-comparison

Edited by G Eysenbach; submitted 27.05.18; peer-reviewed by R David, B Hattink, E Dove; comments to author 19.06.18; revised version received 22.06.18; accepted 10.07.18; published 24.07.18.

Please cite as:
Bott N, Madero EN, Glenn J, Lange A, Anderson J, Newton D, Brennan A, Buffalo EA, Rentz D, Zola S
Device-Embedded Cameras for Eye Tracking–Based Cognitive Assessment: Validation With Paper-Pencil and Computerized Cognitive Composites
J Med Internet Res 2018;20(7):e11143
URL: http://www.jmir.org/2018/7/e11143/
doi:10.2196/11143
PMID:30042093

Review

The Impact of Patient Online Access to Computerized Medical Records and Services on Type 2 Diabetes: Systematic Review

Freda Mold[1*], BSc (Hons), PhD; Mary Raleigh[2*], BSc, RN, RNT, MSc, DClinPractice; Nouf Sahal Alharbi[3*], BSc, MSc, PhD; Simon de Lusignan[4*], BSc, MB BS, MSc, MD (Res)

[1]School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

[2]Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, London, United Kingdom

[3]Department of Health Sciences, College of Applied Studies & Community Service, King Saud University, Riyadh, Saudi Arabia

[4]Department of Clinical and Experimental Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Freda Mold, BSc (Hons), PhD
School of Health Sciences
Faculty of Health and Medical Sciences
University of Surrey
Duke of Kent Building
Stag Hill Campus
Guildford, GU2 7XH
United Kingdom
Phone: 44 01483 684636
Email: freda.mold@surrey.ac.uk

Abstract

Background: Online access to computerized medical records has the potential to improve convenience, satisfaction, and care for patients, and to facilitate more efficient organization and delivery of care.

Objective: The objective of this review is to explore the use and impact of having online access to computerized medical records and services for patients with type 2 diabetes mellitus in primary care.

Methods: Multiple international databases including Medline, Embase, CINAHL, PsycINFO and the Cochrane Library were searched between 2004 and 2016. No limitations were placed on study design, though we applied detailed inclusion and exclusion criteria to each study. Thematic analysis was used to synthesize the evidence. The Mixed Methods Appraisal Toolkit was used to appraise study quality.

Results: A search identified 917 studies, of which 28 were included. Five themes were identified: (1) disparities in uptake by age, gender, ethnicity, educational attainment, and number of comorbidities, with young men in full-time employment using these services most; (2) improved health outcomes: glycemic control was improved, but blood pressure results were mixed; (3) self-management support from improved self-care and shared management occurred especially soon after diagnosis and when complications emerged. There was a generally positive effect on physician-patient relationships; (4) accessibility: patients valued more convenient access when online access to computerized medical records and services work; and (5) technical challenges, barriers to use, and system features that impacted patient and physician use. The Mixed Methods Appraisal Toolkit rated 3 studies as 100%, 19 studies as 75%, 4 studies as 50%, and 1 study scored only 25%.

Conclusions: Patients valued online access to computerized medical records and services, although in its current state of development it may increase disparities. Online access to computerized medical records appears to be safe and is associated with improved glycemic control, but there was a lack of rigorous evidence in terms of positive health outcomes for other complications, such as blood pressure. Patients remain concerned about how these systems work, the rules, and timeliness of using these systems.

(J Med Internet Res 2018;20(7):e235) doi:10.2196/jmir.7858

KEYWORDS

medical records; online access; online services; medical records systems, computerized; computers; primary care; type 2 diabetes mellitus

Introduction

Worldwide, in 2015, 415 million adults aged 20 to 79 years were estimated to have diabetes; and this figure is expected to rise to 642 million by 2040 [1]. The most common type of diabetes is type 2 (type 2 diabetes mellitus, T2DM) and the number of T2DM patients in the UK is steadily growing [2]. Currently, there are 3.2 million people with T2DM, and by 2025 this figure is estimated to reach 5 million [3,4]. A further 630,000 people are predicted to have undiagnosed T2DM [5]. The impact of T2DM is considerable, with the expenditure for treating this condition—and its complications—currently costing the National Health Service £8.8 billion a year, which is over 8% of its annual budget. This expenditure is expected to rise to £15.1 billion by 2035 [6].

Online access to medical records has the potential to support patient-centered care, to improve convenience for patients, and to improve patient satisfaction. Empowering patients by giving them greater access to their medical records and to link online services may, not only assist in self-management of their conditions, but also facilitate organization and delivery of care [7,8]. However, use of these technologies by patients is also a burden for health care providers and there are concerns about privacy and confidentiality [9,10]. Progress has been made in the US health system [11,12], with organizations such as Kaiser Permanente accruing 2 million members who signed up for online services such as appointment bookings, viewing of test results, and emails [13]. However, progress in this regard has been more limited elsewhere in the world.

National systems provide online patient portals separate from their health providers computerized medical records (CMRs) have not been successful in both France and the UK. The French system, Dossier Medical Personnel, was established in 2004 and is a secure CMR system enabling patients direct access to their personal health records. However, by 2013 only 0.31% of the population had opened an account [14]. The English system, “HealthSpace” [15,16], had similarly limited successes with only 0.13% (2913 of the invited 2,442,215) actually signing up and activating their advanced account [16]. Additionally, health professionals in the UK also remain concerned about security, privacy [17-21], and legal constraints [22] of such systems.

In the UK, policy has changed to one which promotes patient access to their medical records via their primary care provider’s CMR system [23]. This access also includes patient online services such as booking appointments, viewing test results, and ordering of prescription refills (repeat prescriptions) [24]. However, email access, which is often part of the provision of such services, is not currently planned.

The aim of this review is to explore the use and impact of having online access to CMR and services for patients with T2DM in primary care.

The objectives are:

- To identify users and nonusers of patient online access to CMRs and services for adults with T2DM (and their caregivers).
- To identify the impact of patients having online access to their CMRs and services in relation to T2DM health outcomes.
- To describe how patient online access to CMRs and services impacts disease management, health delivery, and service access for patients with T2DM.
- To identify any technical challenges, barriers to use and system features which may impact on patients’ uptake and use of online access to CMRs and services.

In identifying these factors, we intend to enhance knowledge of who, why (for what reasons), and when patients use or do not use online access to CMRs and services to manage their diabetes. This is important if we are to identify potential gaps in new service delivery methods; and critical if we are to design innovative services that bridge gaps in current care and design services which are accessible to all.

Methods

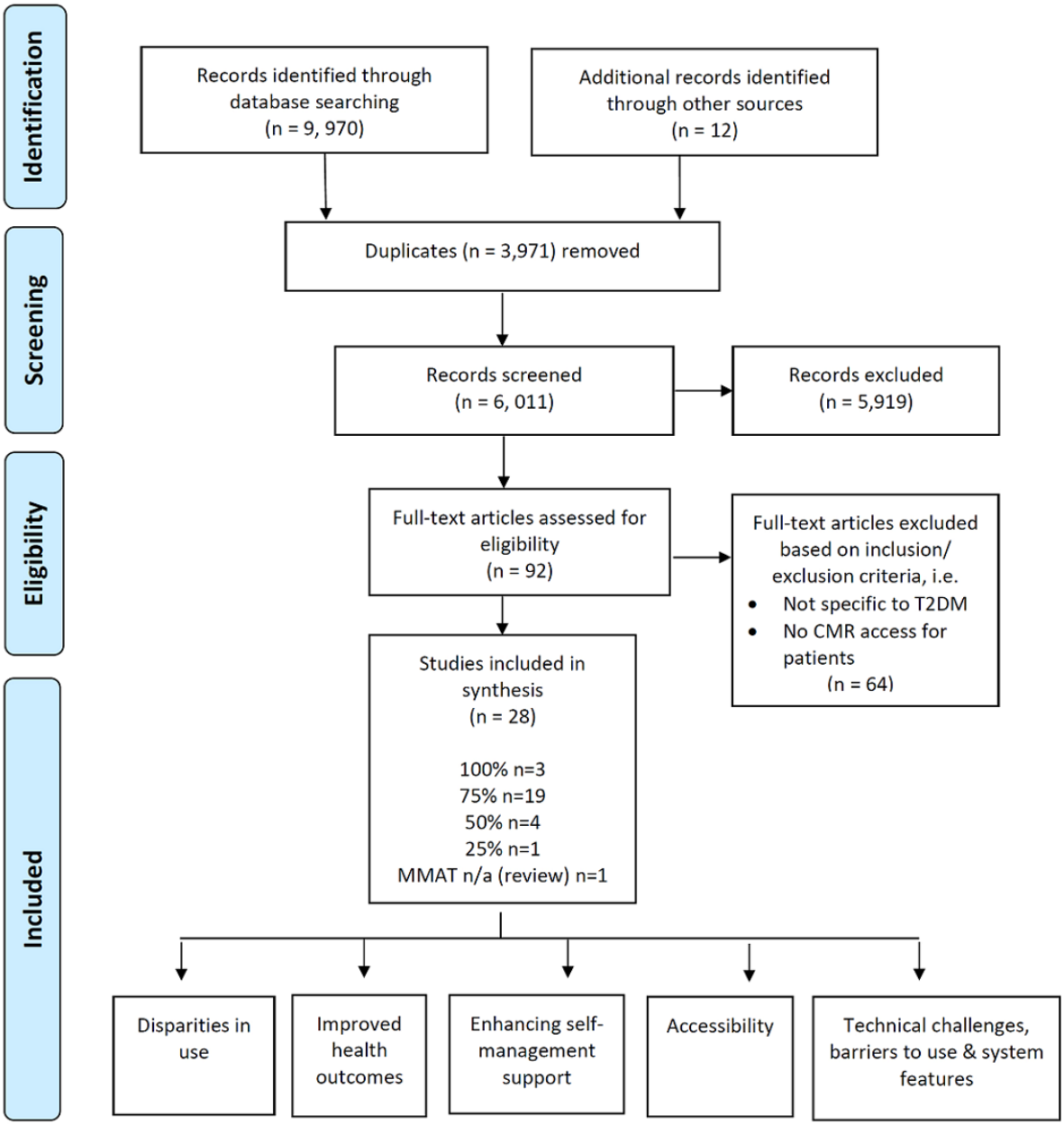
Review Structure

We used a standard methodological approach to conduct a systematic review, as used in our previous studies [25,26]. The evidence sourced in the different stages of this review is displayed using a Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram (Figure 1) [27]. The review aims were structured in a systematic way, using the elements of a clinical question including population, intervention, comparator, and outcome (PICO) [28]. The population (P) included were adults with T2DM and their caregivers, these either being a family member, neighbor, or friend responsible for looking after a person; the intervention (I) was any aspect of online record or service use, the comparator (C) was nonusers of online records or services, and the outcomes (O) were potential impact of online record use or services on the individual (health outcomes), the organization (integration into services), or service technology (current practice information technology [IT] frameworks).

Search Strategy

Generic and disease-specific searches were developed and run across 9 bibliographic databases focusing on online access to CMR and services from 2004 to October 2016. To ensure evidence was as relevant and up-to-date as possible searches were repeated across databases (EBSCO platform) at the end of the review period. The following databases were searched: MEDLINE, Embase, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsycINFO, Cochrane database, Cochrane Effective Practice and Organization of Care Group (EPOC), Database of Abstracts of Reviews of Effects (DARE), and the King’s Fund. A search for unpublished material was conducted using the database OpenGrey. Search strings were tailored to each database according to either Medical Subject Heading (MeSH) or index terms and keywords in the title or abstract. Boolean search functions were used (“AND,” “OR,” and “NOT”). An example MEDLINE search string can be seen in Multimedia Appendix 1.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram used for the systematic review. CMR: computerized medical record, MMAT: Mixed Methods Appraisal Tool, N/A: not applicable, T2DM: type 2 diabetes mellitus.



Inclusion and Exclusion Criteria

Comprehensive inclusion and exclusion criteria were applied to the study, as outlined below.

The inclusion criteria for the found studies were as follows:

- Research focusing on patients and caregivers who have online access to CMRs and online services (which may also include disease-specific portals) via their primary care provider
- Research focusing on patients with T2DM
- Adult patients and their caregiver aged 18 years and over
- All study designs including observational and experimental studies, systematic reviews, and pilot studies which report data.
- Within the date range of 2004-2016

The exclusion criteria for the found studies were as follows:

- When online access to CMR was used by health care staff or researchers only with no patient access
- Studies focusing on the delivery of general health information or education only (information giving) with no online access to CMR by patients
- Studies focusing on the deployment or implementation of new CMR systems in primary care
- Online access to CMR by health care organizations which use data for quality monitoring purposes (ie, Quality and Outcomes Framework [5] only, and do not include any form of patient or carer online access
- A translated copy of article was unavailable
- Research protocols, editorials, or commentary articles were excluded

Screening

The total number of papers identified was 9970, and of these 3971 were duplicate articles. Over six thousand (6011) titles or abstracts were screened by three authors (FM, MR, and NSAH) for articles matching the inclusion and exclusion criteria. After this process, 92 papers remained for inclusion in the review. These papers were subject to full-text review to see if they entirely fulfilled the inclusion criteria. Any disagreement regarding possible inclusion was resolved by discussing the full-text versions. After full-text review, 28 articles were retained for in-depth analysis. The reference lists of these selected articles were also hand searched for other relevant papers matching the eligibility criteria. Search results and the decisions made regarding inclusion or exclusion of each study were stored using Endnote (v7.4).

Data Extraction

A data extraction tool (DEF) was designed by the team to extract relevant information across studies, using Excel. The DEF was initially based on previous designs developed by the first author. The extracted data included the study aims, objectives, population, country of origin, study design, outcomes measures and comparators, methods of analysis used, findings, and study implications. Where possible, all relevant statistical information was also extracted. Data extraction was undertaken independently by two authors (MR and NSH) and checked by FM to ensure consistency and reliability of data being extracted.

Quality Appraisal

Data quality was appraised using the Mixed Methods Appraisal Tool (MMAT), an instrument designed to assess the quality of qualitative, quantitative and mixed methods articles [29,30]. The MMAT has five domains each linked to a specific study design; with each domain containing 4 questions. The MMAT has scaled scoring (ie, 25%, 50%, 75%, and 100%). Each article was appraised independently by an author team member, and disagreements were resolved during team meetings. No articles were excluded on the basis of their MMAT score, but more emphasis is placed on articles weighted at 50% or above. Individual scores are presented in the evidence tables. The interrater reliability of the MMAT score is 0.94 [29]. Two raters appraised each study as above. A final total of 28 articles remained and were subject to full data extraction.

Data Analysis

Thematic analysis was used to identify themes from the evidence. The analysis was guided by the framework offered by Mayring [31]. This method was chosen as it is sensitive to the diverse type of evidence under study, and the large evidence base. A systematic approach was taken throughout, including the analysis in order to minimize any lack of transparency regarding process or analysis decisions. The heterogeneity of the outcomes across the studies made meta-analysis of results impossible. Where necessary, relevant statistical information is provided for each paper; however, this data is not brought together as trial data were not sufficiently homogeneous in terms of primary outcome to provide a meaningful summary.

Results

Study Characteristics

Full data extraction, appraisal, and analysis was conducted on the 28 studies. The majority of the papers originated from the USA (21/28) [32-52], 6 studies were from Europe [53-58], and 1 was Australian [59]. The range of international evidence suggests the international significance of the topic area.

There were a variety of study designs, though the majority employed quantitative methods, using surveys (n=10) [33,35,39,40,45,47,51,54-56] or randomized controlled trials (RCTs, n=5) [34,37,49,52,53]. Several qualitative studies used focus groups and interviews (n=5) [32,36,46,57,59]. Other studies included longitudinal cohort studies (n=3) [38,48,50] and audits (n=3) [41,42,43]. Only one study used a quasi-experimental in design (single interrupted time series) (n=1) [44] and one interpretive review (n=1) [58]. For further information, see Multimedia Appendix 2.

We identified five themes from the studies. These were: (1) disparities in use, (2) improved health outcomes, (3) enhancing self-management support, (4) accessibility, and (5) technical challenges, barriers to use, and system features.

Disparities in Use

We found disparities based on age, level of deprivation, educational status, ethnicity, and differences in people with more comorbidities. There was greater uptake by those participants with higher income, those who reside in more affluent areas, or those with private insurance.

When considering the age of the participants, users with online access to CMRs and services tended to be younger (59 vs 62 years; $P<.01$) [43,54] or in the 50 to 65 years age band [41,56]. One RCT, which explored the use of an e-journal service, reported little difference in the age of enrollees and nonenrollees (48.9 vs 46.7 years; $P<.001$) [40].

Some studies found that online CMR users had a higher mean annual income (US $53,000 vs US $47,500; $P<.01$) [43], they were said to have higher paid jobs [54] and reside in affluent neighborhoods [41]. In contrast, an RCT that explored the use of an e-journal service, reported little difference in the median income between enrollees and nonenrollees (US $54,617 vs US $52,012; $P<.001$) [40]. Insurance status also influenced online service use, with greater uptake of e-journal use in commercially insured users than those privately insured (84.7% vs 74.7%; $P<.0001$) [40].

Online access to CMR and services was generally reported to be greatest for younger males [41,50,56]. One study suggested women over the age of 65 years were less likely to access services compared to men, who were reported to be more familiar with the internet through employment [41] but one RCT reported little difference in previsit e-journal use by gender at enrolment [40].

Patients who use, or request a log-in, for online CMR access and services were also likely to have a higher level of educational attainment [43,55]. Patients without a university degree (compared to college graduates; odds ratio [OR] 2.3,

95% CI 1.9-2.7) were less likely to log on to online CMRs or services [39].

People with T2DM and with multiple comorbidities and polypharmacy were perceived to have greater diabetes-related stress. These patients were more likely to request access to their CMRs [56]. Additionally, a later survey found greater use of a Web-based portal (related to medicated T2DM patients) by patients experiencing more hypo- and hyperglycemic episodes [54]. A retrospective evaluation study also found the use of shared medical records was greater in patients with higher levels of clinical morbidity [41]. Compared to moderate or lower morbidity, those with high clinical morbidity had a 30% higher rate of ongoing use (rate ratio 1.30, 95% CI 1.16-1.45; *P*<.001); and individuals with very high morbidity had a 21% higher use (rate ratio 1.21, 95% CI 1.07-1.37; *P*=.003). Initial CMR use was also more likely within 3 months of an increase in morbidity (hazard ratio 1.61, 95% CI 1.28-2.01) [41].

There were large differences in the use and uptake of secure messaging (SM) services by different ethnic groups. Black or Hispanic patient groups were less likely to register and use patient online services [38,42]. Similarly, significant differences were found between ethnic minority groups (87.1%) compared to Caucasian (users 69.8%; *P*<.001) in completing a previsit electronic journal (e-journal) about their T2DM targets [40]. African-American and Latino patients were also found to have had higher odds of never logging on to a patient portal (OR 2.6, 95% CI 2.3-2.9 and OR 2.3, 95% CI 1.0-2.6) [51]. Black minority groups were also the least likely to use online services (OR 0.25, 95% CI 0.10-0.63) and the internet [45]. Patients who accessed and used CMR and services were, therefore, likely to be Caucasian (84 users vs 66 nonusers, *P*<0.01) compared to African-American (11 users vs 28 nonusers), and other minority groups [43]. However, in contrast, another study found T2DM patients were more likely to place a positive value on online services if they were male (OR 5.8, 95% CI 0.7-48.9), were from an ethnic group (OR 2.1, 95% CI 0.3-17.6) or had been diagnosed with diabetes within the last 5 years (OR 6.0, 95% CI 0.7-49.8) [33].

A cross-sectional survey also found that ethnicity was a significant predictor of shared medical record (SMR) use. Black (34%, 36/107; OR 0.18, 95% CI 0.11-0.30) and Asian (37%, 35/96; OR 0.40, 95% CI 0.20-0.77) T2DM patients were less likely than Caucasian patients to use SMRs (62%, 265/426; *P*<.01) [45].

Health literacy was also found to play a significant role in the use and uptake of online access to CMR and services. A survey of 14,102 T2DM patients reported that those with limited health literacy were less likely to access a portal than those with adequate health literacy [39]. Of the respondents with limited health literacy, 40% (5671) had higher odds of never signing on to a portal (OR 1.7, 95% CI 1.4-1.9) compared with those who were health literate [39].

Frequency and intensity of CMR access and services were also found to be associated with better diabetes knowledge [54]. Frequency and intensity of service use, such as portal access, could also be associated with different types of health users, for example active or nonactive users [50].

Improved Health Outcomes

There was a positive association between the use of online access to CMR and services and improved glycemic control [35,37,43,47,48] and general health care management [46]. However, results for blood pressure (BP) were uncertain with some studies reporting improvements in BP outcomes [34,42,43] and other studies reporting either no change in BP outcomes [37], limited change of BP results over time [53] or there were too few patients within the study to provide a meaningful comparison of BP risk [52].

Frequent use of SM between the physician and T2DM patients allowed medication regimes to be optimized more quickly between in-person visits and was associated with improved glycemic control. HbA_{1c} levels (7%) were 36% higher in the SM user group (with 12 or more threads of correspondence) compared to non-SM user groups (relative risk [RR] 1.36, 95% CI 1.16-1.58) when compared with nonmessaging group [35]. A retrospective longitudinal study to determine the extent to which SM is associated with better glycemic control, found that frequent use of SM in the first year was of use is likely to achieve glycemic control (HbA_{1c}< 7% and <8%; *P*<.05) [48].

Two further studies found that using Web-based CMR was effective in improving diabetes management [37,43]. A pilot RCT found that HbA_{1c} declined by 0.7%, (*P*=.01; 95% CI 0.2-1.3) an average of 8.2% (7/83) among intervention patients compared to 7.9%, (6/83) with usual care (UC) [37]. However, there was no difference in secondary outcome measures: systolic, diastolic blood pressure, and cholesterol levels between pilot intervention and control groups [37]. Similarly, a retrospective audit of HbA_{1c} levels was 0.29% lower (95% CI –0.35 to –0.23; *P*<.01) after 10 days, compared to nonusers [43].

An RCT comparing clinical outcomes of patients who used a home telemedicine unit (including SM, access to medical record data) to those who receive UC found that intervention group hemoglobin improved compared to UC (0.18%; *P*=.006). Mean systolic and diastolic blood pressure level decreased in the intervention group from 142/71mm Hg to 137/68 mm Hg. The net adjusted reduction for systolic was 3.4 mm Hg (*P*=.001) and for diastolic 1.9 mm Hg (*P*<.001) [34].

Online services such as SM and electronic health reminder letters sent via CMRs also resulted in modest improvements in the management of diabetes care. Greater self-reported use of SM to manage medical appointments were significantly associated with better glycemic control (*P*=−0.29; *P*=0.04) [46]. Automatic electronic health reminder letters (sent via CMR) also showed modest improvement in some diabetes measures, but not all [47]. At the end of 12 months, a CMR letter was effective in achieving compliance targets for testing for HbA_{1c} and low-density lipoprotein (LDL; or 1.24, *P*=.005; or 1.35, *P*=0.03; or 1.48, *P*<.001, respectively). However, these improvements were not sustained with postintervention findings indicating a decline in LDL levels in the following 12 months (0.76, *P*=.003) and in the composite endpoint (or 0.78, *P*=.005) [47]. As such, although the proportion of HbA_{1c} checks improved over a 12-month period, there was an overall gradual

decline in achieving an HbA_{1c}<7.0% at each time point [47]. Further evidence suggests a decline in effectiveness over time. Although results from an RCT showed an initial significant decline in HbA_{1c} (0.2%) (*P*=.029) systolic (*P*=.036) and diastolic BP (*P*=.035); there were minimal differences between the intervention and control group for these outcomes at 6 months [53].

A study to determine whether physicians who communicate with their patients using (SM and telephone calls) provide better care for patients, found the use of SM within Black or Hispanic groups were associated with improved outcome scores in HbA_{1c}, cholesterol and blood pressure (*P*<.01) [42].

Enhancing Self-Management Support

Self-management support interventions included in CMR access and services facilitated shared management [36,52,57], patients sense of preparedness [40,46,52], and communication with their health care providers [42,46], including contact outside of conventional working hours [32,36].

Record access was initially reported to improve T2DM shared management and decision-making (DM) between physicians and patients [52,57]. This was reported to result in patients' greater sense of empowerment [52]. A qualitative study showed that self-management of patients' symptoms also improved with online services, such as access to a diabetes-specific portal [33]. The least-valued function of online services was an electronic information board for patients to share and discuss and answer questions in real time (11/21). In a later focus group, study participants felt more in control of symptoms, valued opportunities to view results, and manage their own medication lists. These patients also received health reminders to monitor personal lifestyle goals in order to remain well [36].

Patients who had online access to CMRs and services were also found to be more prepared for upcoming appointments and were more likely to have medication reviews [46,52]. An RCT found online access to CMR enabled them to forward plan for upcoming appointments ensuring adjustments to treatment regimens (53%, n=82 vs 15%, n=41; *P*<.001) when compared to a control group [52]. Another RCT post-intervention survey to measure satisfaction of an e-journal also found that 55.8% (450/806) of patients were better prepared for doctor visits and 58.0% (467/806) providers held more accurate information [40]. Ease of access to consultation information from home (75.5%, 312/413), and opportunities to monitor disease and treatments (42.5%, 132/413) contributed to patients' motivation for requesting a CMR login to monitor their diabetes and treatment [54]. Ease of record access and attitudes towards record ownership were also proxies of service quality [57]

Online access to CMRs and services was also found to improve communication with physicians [42,46] as patients were more satisfied when they could view records, request prescription refills, and have personal control over appointment times [36]. A study describing the experience of patients with a chronic medical condition found they valued online services to communicate with physicians', in comparison to traditional office visits or telephone conversations [32]. Patients also valued seeing results of medical tests online and to track their health status, a need that was previously unmet. Patients felt more secure about managing diabetes symptoms and engaged positively with information provided, especially when the nurse practitioner answered their queries in a timely and consistent manner [32]. Timelines of response was important as users were frustrated when tests results were not released, and messages were not answered [36].

Accessing information outside of normal clinical times was also seen as important [46]. Opportunities for "virtual engagement" outside office hours were reported to potentially reduce demand on providers' time and encourage self-efficacy. Similarly, 62% (13/21) of patients rated SM as a useful way to communicate with community health care teams and services to manage diabetes care [33].

Accessibility: Primarily Using Messaging

SM for T2DM patients via CMRs was associated with higher health care utilization, both in terms of outpatient visits [35] and emergency and primary care contacts [36,44]. However, there were no significant changes reported in the number of patient visits or telephone calls received in primary care; from the implementation of a secure communication system [36] and consultation length was largely unaffected [59].

A cross-sectional study found frequent use of CMR messaging was associated with a higher rate of outpatient visits (RR 1.39, 95% CI 1.26-1.53) and suggested an increase of 3-4 additional visits beyond the normal baseline rate of 9 visits per year [35]. Similarly, a study to test whether SM was associated with increased health care utilization and costs found that as the number primary care visits declined, the level of primary care contact actually increased; largely from the use of SM. This single interrupted time series study to evaluate a new initiative (including SM) found emergency visits increased by 9% annually by full implementation. Annual emergency costs also rose by 13% [44].

An earlier interview study which explored the challenges of implementing a secure eHealth software tool (electronic communication system) found no significant change in the number of patient visits or telephone calls received in the office (preintervention, n=21 and postintervention n=18). However, the frequency of CMR and health reminders views increased; as did SM [36].

A feasibility study to explore controlled online access to CMR between general practitioners and patients using a uniquely tailored USB stick (with patient identifier technology) found minimal impact regarding consultation length [59]. However, this system promoted the accuracy of records by patients being able to view their records and report incorrect entries in their medical records [59].

Finally, a pilot RCT using a shared CMR, found care managers reportedly spending 4 hours per week updating care plans and communicating with patients over the Web; thereby potentially lengthening the working day for some professional groups in primary care [37].

Technical Challenges, Barriers to Use, and System Features

Technical problems with online access to CMR frustrated both patients and providers alike. The consequences were feelings of "disillusionment" with the system and a sense of being "cut off" [32]. Other technical challenges involved lost or unknown passwords and problems with the technical aspects of portals [36]. Other barriers to CMR access were based on expectations as to how online access should work [43] or being unaware of an online portals existence (72.4%, 549/758) [55]. Previous negative experiences and preconceived beliefs or rules about SM were also perceived to be barriers to use [46].

A qualitative study that described the experiences of patients' use of a disease management program (including CMR access and services) found several recurring themes which may impact on the design and use of Web-based tools for T2DM patient groups [32]. Participants expressed how much they appreciated support in managing nonacute concerns and valuing individual communication at convenient times [32]. Patients desire for individual communication could also potentially be important for patients at specific time points, such as for the newly diagnosed. Being able to upload information about blood glucose with a nurse practitioner also provided participants with a "virtual presence." Access to real-time health information and timely feedback on medical tests reduced individual worries, which ultimately facilitated better symptom management [32].

Table 1 shows the review article by study design and research focus. Table 2 reports findings by their respective themes. Multimedia Appendices 3 and 4 present a detailed copy of the evidence tables, outlining key points across all references.

Quality Appraisal Findings

All original studies were subject to MMAT assessment (n=27). The mean MMAT score of included studies was 72% (SD 16.7); indicating moderate to good study quality. Of the 27 included studies, 3 studies were rated as 100% [34,43,49], 19 were rated as 75% [35-37,39-47,50-54,56,57,59], 4 studies were rated as 50% [32,38,48,55], and only 1 study was rated as 25% [33]. See Multimedia Appendix 5, the MMAT Assessment Table for further information.

The majority of the included studies were of moderate to good quality. However, key information relating to outcome measures and comparator groups was occasionally incompletely reported, and some studies lacked detail regarding the description (or processes) of data analysis [33,38,48,56]. MMAT appraisal is useful moving forward as it provides a basis through which to ensure key information is considered at all stages of future research design and reporting.

Table 1. Study design, research focus and Mixed Methods Appraisal Tool (MMAT) score of included studies.

Reference	MMAT score	Study design	Study or intervention aim
Ralston et al 2004 [32]	50	Qualitative study using semistructured interviews	To explore the experiences of diabetes management with CMRs[a] use
Hess et al 2006 [33]	25	Survey and focus group follow up interviews	To evaluate a CMR portal with customized portal features
Shea et al 2006 [34]	100	RCT[b]	To evaluate impact of home telemedicine unit to usual care, on clinical outcomes
Harris et al 2009 [35]	75	Cross-sectional survey	To determine if CMR use is linked to higher quality of care and lower outpatient utilization
Hess et al 2007 [36]	75	Focus groups pre- and postimplementation	To assess patient reaction and challenges with eHealth technology
Ralston et al 2009 [37]	75	Pilot RCT	To test Web-based care management of glycemic control using CMRs
Roblin et al 2009 [38]	50	Longitudinal cohort survey and clustered randomized design	To assess racial preference for registering with a Kaiser Permanente CMR system
Sarkar et al 2010 [39]	75	Survey	Compare use of portal for English-speaking patients versus patients with limited health literacy
Wald et al 2010 [40]	75	RCT-survey	To describe patients experiences of previsit e-Journal use
Weppner et al 2010 [41]	75	Retrospective cohort study	To evaluate the use of SMR[c] between older patients and provider
Bredfeldt et al 2011 [42]	75	Retrospective study	To determine the relationship between effectiveness SM[d] or phone calls and Diabetes Recognition Program scores
Tenforde et al 2011 [43]	100	Retrospective audit	To measure the association of CMR use per days and diabetes quality measures
Grembowski et al 2012 [44]	75	Single interrupted time series-design	To examine whether a Group Health Co-operative changed utilization and cost of care
Lyles et al 2012 [45]	75	Cross-sectional survey	To assess the relationship between race or ethnicity and CMR use
Wade-Vuturo 2013 [46]	75	Mixed methods plus focus groups and survey	To explore how adults with T2DM[e] use a patient portal, to understand nonusers perspectives; and the relationship between SM and glycemic control
Berryman et al 2013 [47]	75	Cross-sectional, practice level study	To evaluate differences in decision making quality metrics at four time points, before and after the introduction of CMR reminders
Harris et al 2013 [48]	50	Retrospective longitudinal cohort plus observational analysis	To determine differences in glycemic control and adherence to HbA_{1c}[f] testing associated with SM
Tang et al 2013 [49]	100	Two-armed RCT. Online questionnaire	To evaluate an online disease management system, compared with usual car
Jones et al 2015 [50]	75	Longitudinal cohort	To describe the types and patterns of portal users in an integrated delivery system
Sarkar et al 2011 [51]	75	Survey	To examine whether social factors influence the use of a patient portal.
Grant et al 2008 [52]	75	RCT	To evaluate the impact of online access to CMR to tailor decision making support and for patient to “develop a plan of care”
Holbrook et al 2009 [53]	75	RCT	To assess the effectiveness of a shared decision support system to improve diabetes care processes & clinical markers
Ronda et al 2015 [54]	75	Survey	To examine patient experiences and use of a Web-portal to access CMR to determine the need for portal redesign
Ronda et al 2014 [55]	50	Cross sectional design/survey	To identify perceived barriers of a Web-based portal to optimize use
Ronda et al 2013 [56]	75	Survey	To examine differences and satisfaction rates of T1DM[g] and T2DM users or nonusers of a web portal
Fisher et al 2009 [57]	75	Focus groups and telephone interviews	To explore patients’ use of CMR, its benefits, impact, and risks

Reference	MMAT score	Study design	Study or intervention aim
Jilka et al 2015 [58]	N/A[h]	Interpretative review	To evaluate the impact of a Patient accessible electronic health records for patients to manage personal clinical information
Bomba et al 2004 [59]	75	Feasibility study with field trial and focus groups	To test the feasibility of building a CMR for access using a USB stick (with unique identifier technology). To evaluate USB access

[a]CMR: computerized medical records.

[b]RCT: randomized controlled trial.

[c]SMR: shared medical record.

[d]SM: secure messaging.

[e]T2DM: type 2 diabetes mellitus;

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

[g]T1DM: type 1 diabetes mellitus.

[h]N/A: not applicable.

Table 2. Themes identified across the included studies.

Reference	Theme 1: Disparities in use	Theme 2: Improved health outcomes	Theme 3: Enhancing self-management support	Theme 4: Accessibility primarily using messaging	Theme 5: Technical challenges, barriers to use, and system features
Ralston et al 2004 [32]			✓		✓
Hess et al 2006 [33]	✓		✓		
Shea et al 2006 [34]		✓			
Harris et al 2009 [35]		✓		✓	
Hess et al 2007 [36]			✓	✓	✓
Ralston et al 2009 [37]		✓		✓	
Roblin et al 2009 [38]	✓				
Sarkar et al 2010 [39]	✓				
Wald et al 2010 [40]	✓		✓		
Weppner et al 2010 [41]	✓				
Bredfeldt et al 2011 [42]	✓	✓	✓		
Tenforde et al 2011 [43]	✓	✓			✓
Grembowski et al 2012 [44]				✓	
Lyles et al 2012 [45]	✓				
Wade-Vuturo 2013 [46]		✓	✓		✓
Berryman et al 2013 [47]		✓			
Harris et al 2013 [48]		✓			
Tang et al 2013 [49]					
Jones et al 2015 [50]	✓				
Sarkar et al 2011 [51]	✓				
Grant et al 2008 [52]		✓	✓		
Holbrook et al 2009 [53]		✓			
Ronda et al 2015 [54]	✓				
Ronda et al 2014 [55]	✓				✓
Ronda et al 2013 [56]	✓				
Fisher et al 2009 [57]			✓		
Jilka et al 2015 [58]					
Bomba et al 2004 [59]				✓	

Discussion

Principal Results

Online access appears to be valued by patients with T2DM [32,36,40] but in its current state of development it may widen disparities [39,41,45,55,56]. Males in full-time employment with good IT skills are those most likely to use this service [41]. There appears to be little provision, or development of systems to meet the needs of caregivers; who often provide support outside of working hours.

There are also differences in online access to CMR and services between ethnic groups [38,39,40,42,45]. Black and Asian ethnic groups [45], and Hispanic [42] and African-American male patients [38] were less likely to register and use online services [45], including portals [38,39,40,42]. Only one study suggested gender differences in online access to CMR for African-American patients [38]. Further evidence is needed to explore this area.

Online access to CMR and services is much greater soon after diagnosis, when needs become complex and where changes are needed in medication [32,41,54-56]. Suggesting use could be of benefit to patients at specific time points in their care.

People who take up online services have better glycemic control [35,37,43,48]. However, to date, there is limited evidence of improved outcomes, in either macro- or microvascular complications. Other outcomes such as blood pressure had mixed results either reporting a decline in BP [34,42,43], no change in BP [37], or study limitations which impacted on BP reporting [52,53].

Patients remain concerned about specific aspects of online access to CMR and services including residual worries about how these systems work [43], the rules of engagement in using these systems [46], timeliness of responses from health care professionals [36], and technical failures [32].

Implications for Future Practice and Research

This review shows disparities between patient groups' online access to CMR and services to manage diabetes. Greater efforts are needed to make these technologies available to a wider group of patients. This includes across ethnic groups, patients with varying levels of information technology and literacy skills, and age groups. Codesign processes may help identify and meet the needs of patients and caregivers, as their insights may bridge gaps in these new service delivery systems. Further research is needed to understand more about who, why (for what reasons), and when patients use or do not use online access to CMR and services to manage their diabetes.

Online access to CMR and services may need to be tailored to the specific user and condition. This may be particularly important for acute complications for example ketoacidosis. Caregivers may also have different requirements depending on the care recipients specific condition, comorbidities, and wishes about sharing their medical data.

Evidence suggests greater uptake at the time of diagnosis and for a period after, but use does not persist [56]. Further research is needed to explore why use of CMR drops away in the period following initial diagnosis.

Research into physicians and patients views about CMR access in terms of how to provide caregivers appropriate access privileges has not been fully addressed. Whilst physicians are rightly concerned about privacy and confidentiality [58], patients' concerns focused more specifically on functionality, technical support, and system knowledge [32,36,43,46,55]. It could be that the data needed for monitoring and care in diabetes should have a different level of access, without allowing caregivers comprehensive access to a patients' record. This might allow sharing of diabetes management with caregivers, with the patient's consent, without making all their health information available.

Future research should continue to study and address health literacy [38,39,43] and ethnic differences in patients' access [38,39,40,42,45]. Potential language barriers and lack of explanation of medical terms may also contribute to unequal access [54]. Further research should also be mindful of any unanticipated consequences of online service use in terms of unequal access and use [38].

Online access to CMR and services has also shown to impact on patients' self-care behaviors which may influence the physician-patient relationship [54,57,60]. It would be interesting to assess in what ways these revised styles of communication impact on service use and/or uptake.

Information technology systems supporting online access to CMR require future development in order to engage and sustain physician and patient use [52]. Tailoring online services to disease-specific conditions may be seen as a valuable resource both in terms of care delivery [33,41,51] and in relation to self-care [33].

Improvements to online access to CMR and services designs may support bundles of care for T2DM management [53] or to improve poorly controlled diabetes [49]. Patient online services could allow targeted approaches to engaging with different population groups with incentives and messages to motivate technology use [50]. However, improving access will be challenging unless there is adequate future funding and training [34].

Integration into primary care business process can be challenging and these include data management [61], communication [42] and costs of implementation and sustainability [44]. Whilst integrating Web-based technology into primary care has been relatively easy [62], health care professionals, may not quickly change their communication patterns [36].

Deployment of online medical records globally is gathering pace [60,63,64]. Within the UK, the importance of online access to CMR and services is growing; as demand for primary care coverage to be available out-of-core working hours (8 am to 6.30 pm Monday to Friday) [65,66] and in response to service needs to support people in the community [67,68].

There are different models of health care delivery and cost, compared to the UK's National Health Service. Differences

may emerge in the use, design, and adoption of online access to CMRs and services. There is a dearth of evidence emerging from the operation of many national CMR systems such as Australia's "My Health Record System," Hong Kong's "Electronic Health record Sharing System" and others in the United States [69].

Limitations

Like all reviews, evidence has been gathered from various resources from a specific time period. As such there may be several newly published studies that have not been included in this review. Another limitation was the quality of the studies varied (such as poor or incomplete reporting of the study). Findings from the MMAT appraisal indicates possible areas of further development in the design and reporting of studies; particularly in relation to key information such as outcome measures, comparator group data, and description of the data analysis.

All studies reviewed originate from the USA, Australia, and Europe, with little from Africa, Asia, or South America. Limited translation of evidence may have contributed to this lack of evidence. In adhering to the review process, however, every attempt was made to include international evidence which met the inclusion criteria.

Conclusions

Evidence reported in this review show there are disparities in how different patient groups view, access and use these systems to manage their T2DM. Current users of online CMR access and services tend to be young employed men and they are used less by ethnic minority groups. Uptake is also greater after diagnosis, but then usage falls away, and we are not sure why. Online access is used more where there are complex needs or when medication regimens change. Online access in T2DM is associated with improved glycemic control, but as yet there is no clear evidence of improved outcomes in terms of other complications; such as BP. Concerns remain for patients and physicians about the use and integration of these systems. Further research is ultimately needed into how these systems can meet the needs of wider patient groups. Patient online access to CMR and services to support patients with T2DM are well established internationally and are here to stay.

Acknowledgments

The original systematic review study was supported by the Royal College of General Practitioners and commissioned by the Department of Health. This additional work was unfunded.

Authors' Contributions

FM was involved in all aspects of this paper from review focus/design, search string development, sourcing and screening of papers, extracting data, undertaking data analysis and writing of the paper. FM also coordinated and addressed manuscript comments. MR was involved in extracted data, data analysis, writing of this paper, and addressing manuscript comments. NSA was involved in data extraction, the writing of this paper, and addressing manuscript comments. SdL was involved in critique and redrafting versions of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Medline string.

[PDF File (Adobe PDF File), 29KB - jmir_v20i7e235_app1.pdf]

Multimedia Appendix 2

Study characteristics table.

[PDF File (Adobe PDF File), 28KB - jmir_v20i7e235_app2.pdf]

Multimedia Appendix 3

Evidence table: online computerized medical records and services for patients with type 2 diabetes mellitus.

[PDF File (Adobe PDF File), 58KB - jmir_v20i7e235_app3.pdf]

Multimedia Appendix 4

Continued evidence table: online computerized medical records and services for patients with type 2 diabetes mellitus.

[PDF File (Adobe PDF File), 784KB - jmir_v20i7e235_app4.pdf]

Multimedia Appendix 5

Mixed Methods Appraisal Tool (MMAT).

[PDF File (Adobe PDF File), 55KB - jmir_v20i7e235_app5.pdf]

References

1. International Diabetes Federation. Belgium: The International Diabetes Federation; 2013. Prevention of Type 2 Diabetes URL: https://www.idf.org/54-our-activities/225-prevention-of-type-2-diabetes.html [accessed 2018-06-14] [WebCite Cache ID 70AmsVAyN]
2. UK Government. London, UK: Public Health England; 2015. A systematic review and meta-analysis assessing the effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes mellitus in routine practice URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/456147/PHE_Evidence_Review_of_diabetes_prevention_programmes-_FINAL.pdf [accessed 2018-06-14] [WebCite Cache ID 70Am51MiH]
3. Diabetes UK. 2014. Diabetes: Facts and stats URL: https://www.diabetes.org.uk/professionals/position-statements-reports/statistics [accessed 2018-06-14] [WebCite Cache ID 70AmTlJgo]
4. Holman N, Forouhi NG, Goyder E, Wild SH. The Association of Public Health Observatories (APHO) Diabetes Prevalence Model: estimates of total diabetes prevalence for England, 2010-2030. Diabet Med 2011 May;28(5):575-582. [doi: 10.1111/j.1464-5491.2010.03216.x] [Medline: 21480968]
5. Prescribing and Primary Care Team Health and Social Care Information Centre. National Archives UK. London, UK: Health and Social Care Information Centre; 2013. Quality and Outcomes Framework (QOF) 2012/3 URL: http://webarchive.nationalarchives.gov.uk/20180328130852tf_/http://content.digital.nhs.uk/catalogue/PUB12262/qual-outc-fram-12-13-rep.pdf/ [accessed 2018-06-14] [WebCite Cache ID 70Amjm4SW]
6. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. Diabet Med 2012 Jul;29(7):855-862. [doi: 10.1111/j.1464-5491.2012.03698.x] [Medline: 22537247]
7. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Sep 08;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]
8. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]
9. Beard L, Schein R, Morra D, Wilson K, Keelan J. The challenges in making electronic health records accessible to patients. J Am Med Inform Assoc 2012;19(1):116-120 [FREE Full text] [doi: 10.1136/amiajnl-2011-000261] [Medline: 22120207]
10. Cross M. BMA warns against letting patients have access to their electronic records. BMJ 2011 Jan 12;342:d206 [FREE Full text] [Medline: 21228010]
11. Baer D. Patient-physician e-mail communication: the kaiser permanente experience. J Oncol Pract 2011 Jul;7(4):230-233 [FREE Full text] [doi: 10.1200/JOP.2011.000323] [Medline: 22043186]
12. Flynn D, Gregory P, Makki H, Gabbay M. Expectations and experiences of eHealth in primary care: a qualitative practice-based investigation. Int J Med Inform 2009 Sep;78(9):588-604. [doi: 10.1016/j.ijmedinf.2009.03.008] [Medline: 19482542]
13. Silvestre A, Sue VM, Allen JY. If you build it, will they come? The Kaiser Permanente model of online health care. Health Aff (Millwood) 2009;28(2):334-344. [doi: 10.1377/hlthaff.28.2.334] [Medline: 19275988]
14. de LS, Seroussi B. A comparison of English and French approaches to providing patients access to Summary Care Records: scope, consent, cost. Stud Health Technol Inform 2013;186:61-65. [Medline: 23542968]
15. Greenhalgh T, Wood GW, Bratan T, Stramer K, Hinder S. Patients' attitudes to the summary care record and HealthSpace: qualitative study. BMJ 2008 Jun 7;336(7656):1290-1295 [FREE Full text] [doi: 10.1136/bmj.a114] [Medline: 18511764]
16. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010;341:c5814 [FREE Full text] [Medline: 21081595]
17. Honeyman A, Cox B, Fisher B. Potential impacts of patient access to their electronic care records. Inform Prim Care 2005;13(1):55-60 [FREE Full text] [Medline: 15949176]
18. Vodicka E, Mejilla R, Leveille SG, Ralston JD, Darer JD, Delbanco T, et al. Online access to doctors' notes: patient concerns about privacy. J Med Internet Res 2013;15(9):e208 [FREE Full text] [doi: 10.2196/jmir.2670] [Medline: 24072335]
19. Mandl KD, Szolovits P, Kohane IS. Public standards and patients' control: how to keep electronic medical records accessible but private. BMJ 2001 Feb 03;322(7281):283-287 [FREE Full text] [Medline: 11157533]
20. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, et al. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]

21. Car J, Sheikh A. Email consultations in health care: 2--acceptability and safe application. BMJ 2004 Aug 21;329(7463):439-442 [FREE Full text] [doi: 10.1136/bmj.329.7463.439] [Medline: 15321903]
22. Meltsner M. A patient's view of OpenNotes. Ann Intern Med 2012 Oct 02;157(7):523-524. [doi: 10.7326/0003-4819-157-7-201210020-00012] [Medline: 23027321]
23. UK Government. London, UK: The Department of Health; 2012. The power of information: Putting all of us in control of the health and care information we need URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213689/dh_134205.pdf [accessed 2018-06-14] [WebCite Cache ID 70Ao7rYnD]
24. NHS England. 2016. Patient Online URL: https://www.england.nhs.uk/ourwork/pe/patient-online/ [accessed 2018-06-18] [WebCite Cache ID 70GYHi0VQ]
25. Mold F, Ellis B, de Lusignan LS, Sheikh A, Wyatt JC, Cavill M, et al. The provision and impact of online patient access to their electronic health records (EHR) and transactional services on the quality and safety of health care: systematic review protocol. Inform Prim Care 2012;20(4):271-282 [FREE Full text] [Medline: 23890339]
26. Mold F, Ellis B, de Lusignan S, Cavill M, Michalakidis G, Sheikh A, et al. The provision and impact of online patient access to their electronic health records and transactional services on the quality and safety of health care: systematic review protocol. PROSPERO 2012:CRD 3091 Dec;International Prospective Register of Systematic Reviews Registration Number:4201 [FREE Full text]
27. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
28. Stillwell SB, Fineout-Overholt E, Melnyk BM, Williamson KM. Evidence-based practice, step by step: asking the clinical question: a key step in evidence-based practice. Am J Nurs 2010 Mar;110(3):58-61. [doi: 10.1097/01.NAJ.0000368959.11129.79] [Medline: 20179464]
29. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. Int J Nurs Stud 2012 Jan;49(1):47-53. [doi: 10.1016/j.ijnurstu.2011.07.002] [Medline: 21835406]
30. Pluye P, Gagnon M, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
31. Mayring P. Qualitative Content Analysis. Forum: Qualitative Social Research On-line Journal, 1(2) 2000;1(2):1-10 [FREE Full text]
32. Ralston JD, Revere D, Robins LS, Goldberg HI. Patients' experience with a diabetes support programme based on an interactive electronic medical record: qualitative study. BMJ 2004 May 15;328(7449):1159 [FREE Full text] [doi: 10.1136/bmj.328.7449.1159] [Medline: 15142919]
33. Hess R, Bryce CL, McTigue K, Fitzgerald K, Zickmund S, Olshansky E. The diabetes patient portal: Patient perspectives on structure and delivery. Diabetes Spectrum 2006;19(2):106-110 [FREE Full text]
34. Shea S, Weinstock RS, Starren J, Teresi J, Palmas W, Field L, et al. A randomized trial comparing telemedicine case management with usual care in older, ethnically diverse, medically underserved patients with diabetes mellitus. J Am Med Inform Assoc 2006;13(1):40-51 [FREE Full text] [doi: 10.1197/jamia.M1917] [Medline: 16221935]
35. Harris LT, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. Diabetes Care 2009 Jul;32(7):1182-1187 [FREE Full text] [doi: 10.2337/dc08-1771] [Medline: 19366959]
36. Hess R, Bryce CL, Paone S, Fischer G, McTigue KM, Olshansky E, et al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. Telemed J E Health 2007 Oct;13(5):509-517. [doi: 10.1089/tmj.2006.0089] [Medline: 17999613]
37. Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care 2009 Feb;32(2):234-239 [FREE Full text] [doi: 10.2337/dc08-1220] [Medline: 19017773]
38. 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]
39. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
40. Wald JS, Businger A, Gandhi TK, Grant RW, Poon EG, Schnipper JL, et al. Implementing practice-linked pre-visit electronic journals in primary care: patient and physician use and satisfaction. J Am Med Inform Assoc 2010;17(5):502-506 [FREE Full text] [doi: 10.1136/jamia.2009.001362] [Medline: 20819852]

41. Weppner WG, Ralston JD, Koepsell TD, Grothaus LC, Reid RJ, Jordan L, et al. Use of a shared medical record with secure messaging by older patients with diabetes. Diabetes Care 2010 Nov;33(11):2314-2319 [FREE Full text] [doi: 10.2337/dc10-1124] [Medline: 20739686]
42. Bredfeldt CE, Compton-Phillips AL, Snyder MH. Effects of between visit physician-patient communication on Diabetes Recognition Program scores. Int J Qual Health Care 2011 Dec;23(6):664-673 [FREE Full text] [doi: 10.1093/intqhc/mzr061] [Medline: 21937586]
43. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
44. Grembowski D, Anderson ML, Ralston JD, Martin DP, Reid R. Does a large-scale organizational transformation toward patient-centered access change the utilization and costs of care for patients with diabetes? Med Care Res Rev 2012 Oct;69(5):519-539. [doi: 10.1177/1077558712446705] [Medline: 22653416]
45. 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]
46. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
47. Berryman SH, Sick BT, Wang Q, Swan PJ, Weber-Main AM. Use of automated reminder letters to improve diabetes management in primary care: outcomes of a quality improvement initiative. Qual Prim Care 2013;21(6):359-368. [Medline: 24512834]
48. Harris LT, Koepsell TD, Haneuse SJ, Martin DP, Ralston JD. Glycemic control associated with secure patient-provider messaging within a shared electronic medical record: a longitudinal analysis. Diabetes Care 2013 Sep;36(9):2726-2733 [FREE Full text] [doi: 10.2337/dc12-2003] [Medline: 23628618]
49. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 1;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
50. Jones JB, Weiner JP, Shah NR, Stewart WF. The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res 2015 Feb 20;17(2):e42 [FREE Full text] [doi: 10.2196/jmir.3157] [Medline: 25707036]
51. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
52. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2008 Sep 8;168(16):1776-1782 [FREE Full text] [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]
53. Holbrook A, Thabane L, Keshavjee K, Dolovich L, Bernstein B, Chan D, et al. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. CMAJ 2009 Jul 7;181(1-2):37-44 [FREE Full text] [doi: 10.1503/cmaj.081272] [Medline: 19581618]
54. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Patients' Experiences with and Attitudes towards a Diabetes Patient Web Portal. PLoS One 2015 Jun;10(6):e0129403 [FREE Full text] [doi: 10.1371/journal.pone.0129403] [Medline: 26086272]
55. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
56. Ronda MCM, Dijkhorst-Oei L, Gorter KJ, Beulens JWJ, Rutten GEHM. 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]
57. Fisher B, Bhavnani V, Winfield M. How patients use access to their full health records: a qualitative study of patients in general practice. J R Soc Med 2009 Dec;102(12):539-544 [FREE Full text] [doi: 10.1258/jrsm.2009.090328] [Medline: 19966130]
58. Jilka SR, Callahan R, Sevdalis N, Mayer EK, Darzi A. “Nothing About Me Without Me”: An Interpretative Review of Patient Accessible Electronic Health Records. J Med Internet Res 2015;17(6):e161 [FREE Full text] [doi: 10.2196/jmir.4446] [Medline: 26123476]
59. Bomba D, Fulcher J, Dalley A. Construction of a diabetes database and pilot evaluation of iKey controlled GP-patient access. The Journal on Information Technology in Healthcare 2004;2(5):329-339 [FREE Full text]
60. Jobanputra K, Greig J, Shankar G, Perakslis E, Kremer R, Achar J, et al. Electronic medical records in humanitarian emergencies - the development of an Ebola clinical information and patient management system. F1000Res 2016;5:1477 [FREE Full text] [doi: 10.12688/f1000research.8287.3] [Medline: 28357037]
61. Yau GL, Williams AS, Brown JB. Family physicians' perspectives on personal health records: qualitative study. Can Fam Physician 2011 May;57(5):e178-e184 [FREE Full text] [Medline: 21642732]

62. Royal College of General Practitioners. London, UK: Royal College of General Practitiioners; 2013. A Vision for General Practice in the Future NHS: The 2022 GP URL: http://www.rcgp.org.uk/policy/rcgp-policy-areas/general-practice-2022.aspx [accessed 2018-06-18] [WebCite Cache ID 70GqR40V0]
63. Dainton C, Chu CH. A review of electronic medical record keeping on mobile medical service trips in austere settings. Int J Med Inform 2017 Feb;98:33-40. [doi: 10.1016/j.ijmedinf.2016.11.008] [Medline: 28034410]
64. Jawhari B, Keenan L, Zakus D, Ludwick D, Isaac A, Saleh A, et al. Barriers and facilitators to Electronic Medical Record (EMR) use in an urban slum. Int J Med Inform 2016 Oct;94:246-254 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.07.015] [Medline: 27573333]
65. Baker M, Thomas M, Mawby R. London, UK: Royal College of General Practitioners; 2014. The Future of GP Out-of-Hours Car. A RCGP Position Statement URL: http://www.rcgp.org.uk/policy/rcgp-policy-areas/out-of-hours.aspx [accessed 2018-06-20] [WebCite Cache ID 70Jyzq7cV]
66. Ware J, Mawby R. London, UK: Royal College of General Practitioners; 2015. Patient access to general practice: ideas and challenges from the front line URL: http://www.rcgp.org.uk/policy/rcgp-policy-areas/access-to-general-practice.aspx [accessed 2017-10-12] [WebCite Cache ID 70JzDmXNV]
67. Department of Health. London, UK: Department of Health (Public Health Policy & Strategy Unit/NHS Commissioning Unit); 2013. Living Well for Longer: A call to action to reduce avoidable premature mortality URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/181103/Living_well_for_longer.pdf [accessed 2018-06-18] [WebCite Cache ID 70Gi1EsyO]
68. Department of Health. London, UK: Department of Health, Public Health Policy and Strategy Unit; 2015 Mar 01. Living Well for Longer: One year on URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/416442/Living_Well_for_Longer_-_One_year_on__March_2015_.pdf [accessed 2018-06-18] [WebCite Cache ID 70GhqMK0a]
69. Bowden T, Coiera E. The role and benefits of accessing primary care patient records during unscheduled care: a systematic review. BMC Med Inform Decis Mak 2017 Sep 22;17(1):138 [FREE Full text] [doi: 10.1186/s12911-017-0523-4] [Medline: 28938900]

Abbreviations

BP: blood pressure
CINAHL: Cumulative Index to Nursing and Allied Health Literature
CMR: computerized medical record
DARE: Database of Abstracts of Reviews of Effects
EPOC: Effective Practice and Organization of Care Group
HbA_{1c}: glycated hemoglobin
LDL: low-density lipoprotein
MeSH: Medical Subject Heading
MMAT: Mixed Methods Appraisal Tool
OR: odds ratio
PICO: population, intervention, comparator, and outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
RR: relative risk
SM: secure messaging
SMR: shared medical record
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 04.05.17; peer-reviewed by Y Wang, P Seth, J St-Maurice; comments to author 25.07.17; revised version received 28.11.17; accepted 15.05.18; published 06.07.18.

Please cite as:
Mold F, Raleigh M, Alharbi NS, de Lusignan S
The Impact of Patient Online Access to Computerized Medical Records and Services on Type 2 Diabetes: Systematic Review
J Med Internet Res 2018;20(7):e235
URL: http://www.jmir.org/2018/7/e235/
doi:10.2196/jmir.7858
PMID:29980499

Original Paper

Unsupervised Machine Learning to Identify High Likelihood of Dementia in Population-Based Surveys: Development and Validation Study

Laurent Cleret de Langavant[1,2,3,4,5], MD, PhD; Eleonore Bayen[5,6,7,8], MD, PhD; Kristine Yaffe[5,9], MD

[1]Service de Neurologie, Hôpital Henri Mondor, Assistance Publique Hôpitaux de Paris, Créteil, France

[2]Département de Neurologie, Faculté de Médecine, Université de Paris Est, Créteil, France

[3]Neuropsychologie Interventionnelle U955 EQ1, Institut Mondor de Recherche Biomedicale, Inserm, Créteil, France

[4]Département d'Études Cognitives, École Normale Supérieure, PSL Research University, Paris, France

[5]Global Brain Health Institute, Memory and Aging Center, University of California, San Francisco, CA, United States

[6]Service de Médecine Physique et Réadaptation, Hôpital de la Pitié Salpêtrière, Assistance Publique Hôpitaux de Paris, Paris, France

[7]Médecine Physique et Réadaptation, Faculté de Médecine, Sorbonne University, Paris, France

[8]Laboratoire d'Économie et de Gestion des Organisations de Santé, Département d'Économie de la Santé, Université Paris-Dauphine, Paris, France

[9]Center for Population Brain Health, Department of Psychiatry, Neurology and Epidemiology and Biostatistics, University of California, San Francisco, CA, United States

Corresponding Author:
Laurent Cleret de Langavant, MD, PhD
Service de Neurologie
Hôpital Henri Mondor
Assistance Publique Hôpitaux de Paris
51 Avenue du Maréchal de Lattre de Tassigny
3ème étage
Créteil, 94000
France
Phone: 33 149812308
Fax: 33 149812326
Email: laurent.cleret@gbhi.org

Abstract

Background: Dementia is increasing in prevalence worldwide, yet frequently remains undiagnosed, especially in low- and middle-income countries. Population-based surveys represent an underinvestigated source to identify individuals at risk of dementia.

Objective: The aim is to identify participants with high likelihood of dementia in population-based surveys without the need of the clinical diagnosis of dementia in a subsample.

Methods: Unsupervised machine learning classification (hierarchical clustering on principal components) was developed in the Health and Retirement Study (HRS; 2002-2003, N=18,165 individuals) and validated in the Survey of Health, Ageing and Retirement in Europe (SHARE; 2010-2012, N=58,202 individuals).

Results: Unsupervised machine learning classification identified three clusters in HRS: cluster 1 (n=12,231) without any functional or motor limitations, cluster 2 (N=4841) with walking/climbing limitations, and cluster 3 (N=1093) with both functional and walking/climbing limitations. Comparison of cluster 3 with previously published predicted probabilities of dementia in HRS showed that it identified high likelihood of dementia (probability of dementia >0.95; area under the curve [AUC]=0.91). Removing either cognitive or both cognitive and behavioral measures did not impede accurate classification (AUC=0.91 and AUC=0.90, respectively). Three clusters with similar profiles were identified in SHARE (cluster 1: n=40,223; cluster 2: n=15,644; cluster 3: n=2335). Survival rate of participants from cluster 3 reached 39.2% (n=665 deceased) in HRS and 62.2% (n=811 deceased) in SHARE after a 3.9-year follow-up. Surviving participants from cluster 3 in both cohorts worsened their functional and mobility performance over the same period.

Conclusions: Unsupervised machine learning identifies high likelihood of dementia in population-based surveys, even without cognitive and behavioral measures and without the need of clinical diagnosis of dementia in a subsample of the population. This method could be used to tackle the global challenge of dementia.

(J Med Internet Res 2018;20(7):e10493) doi:10.2196/10493

KEYWORDS

dementia; cognition disorders; health surveys; electronic health records; diagnosis; unsupervised machine learning; cluster analysis; data mining

Introduction

The number of cases of dementia is projected to triple by 2050 worldwide, with a steeper increase in Latin America and the Caribbean, Africa, Asia, and Oceania [1]. Up to half of older adults with dementia are not diagnosed in high-income countries [2] and this proportion is thought to be even higher in low- and middle-income countries [3]. For example, 77% of people with dementia may be underdiagnosed in Brazil [4]. Not only is the diagnosis of dementia complex because it usually relies on an extensive evaluation, but it is also costly [5]. As a consequence, epidemiological data related to dementia mainly comes from Western Europe, East Asia, and North America, but it remains scarce for other regions [6].

Population-based surveys of aging may represent an underinvestigated source of information to study dementia and its determinants. In population-based surveys, nonmedical interviewers collect a rich set of sociodemographic, health, and functional information from nonclinical and representative populations. Several population surveys across four continents are modeled according to the same protocol as the Health and Retirement Study (HRS) in the United States, providing the opportunity to compare aging outcomes between different countries [7]. However, with a few exceptions, dementia outcome is not reliably reported in these surveys.

Machine learning algorithms can assess vast numbers of variables in large datasets, looking for combinations to reliably predict outcomes [8]. This is the case for supervised machine learning algorithms, which learn from specific outcomes available in a subset of individuals, such as the clinical diagnosis of dementia, to expand the acquired knowledge to the whole sample, thanks to a statistical model. In a different approach, unsupervised machine learning algorithms correspond to data-driven techniques that automatically learn from the relationships between elementary bits of information associated with each variable of a dataset. Contrary to supervised machine learning, unsupervised machine learning algorithms unbiasedly reveal associations or clusters existing within datasets without any a priori teaching model.

Because any representative sample of an aging population includes a subgroup of persons living with dementia, we expect unsupervised machine learning to automatically discover this subgroup, without the need of the clinical diagnosis of dementia in a subsample of this population. Our objective was to develop a methodology capable to identify individuals with high likelihood of dementia using a specific unsupervised machine learning algorithm, hierarchical clustering on principal components, in a development cohort in the United States—the HRS [9]. We tested the accuracy of this method by comparing the outcome classification with the predicted probabilities of dementia according to a previously computed supervised machine learning model [10] using the same HRS dataset and based on the clinical diagnosis of dementia available for a subset of the HRS cohort in the Aging, Demographics, and Memory Study (ADAMS) [11]. We also explored the impact of removing cognitive and behavioral measures from the classification algorithm. We then applied this methodology to a validation cohort, the Survey of Health, Ageing and Retirement in Europe (SHARE) [12]. In both cohorts, longitudinal follow-up during the two waves following classification was used as an additional method to validate the clinical relevance of the unsupervised machine learning method.

Methods

Populations

The HRS is a longitudinal population-based survey exploring the health and economic well-being of adults older than 50 years in the United States done every two years since 1992. The HRS sample is stratified geographically and covers all demographic groups. The respondent is randomly selected from all age-eligible household members. Although baseline interviews are conducted face-to-face, follow-up interviews are done by telephone (until 2004), with the exception of respondents older than age 80 years [9]. We chose cross-sectional data from wave 6 of the HRS (January 2002 to February 2003; 18,165 individuals) as our development cohort.

The SHARE is a longitudinal population-based survey of individuals aged 50 years or older based on the same protocol as the HRS. We chose wave 4 of SHARE (May 2010 to April 2012; 58,202 individuals from 16 countries including Austria, Belgium, Czech Republic, Denmark, Estonia, France, Germany, Hungary, Italy, Netherlands, Poland, Portugal, Slovenia, Spain, Sweden, and Switzerland) as our validation cohort.

Measures

Harmonized data (ie, with identically defined variables), from both the HRS and SHARE surveys were downloaded from the Gateway to Global Aging platform [7]. We included variables covering the domains of demographics, health, health care utilization, and cognition that were present in both the HRS and SHARE cohorts. Variables related to insurance, income, financial and housing wealth were removed a priori from both cohort datasets in order to develop a classification method applicable whatever the economic context. Two different behavioral scales were used for the evaluation of depression:

The Center for Epidemiological Study of Depression scale (CES-D) [13] for HRS and the European Union initiative to compare symptoms of depression scale (EURO-D) [14] for SHARE. A total of 92 variables were selected for the HRS cohort and 91 for SHARE (Multimedia Appendix 1). Variables with more than 33% of the data missing were arbitrarily discarded. The remaining missing values were imputed with the regularized iterative principal component analysis (PCA) algorithm of the missMDA package of R software [15], which allows the imputed values to have no weight on the PCA results. This imputation method is complementary to the unsupervised machine learning algorithm we used for classification. Three different datasets were used for each cohort: one with the previous selected variables, one omitting cognitive measures, and one omitting both cognitive and behavioral measures.

Unsupervised Machine Learning Classification

We ran an agglomerative hierarchical clustering on the 10 first principal components resulting from PCA of the datasets (FactoMineR package, R software) [16]. Hierarchical clustering on principal components was considered to be the best unsupervised machine learning technique in this context because (1) PCA allows to reduce the number of variables without losing important information, (2) hierarchical clustering is a very stable method compared to other unsupervised machine learning techniques (eg, *k*-means), and (3) it complements the imputation method we used to obtain a dataset without missing values. The optimal number of clusters in each dataset was determined after running the NbClust package of R software [17].

Predicted Probability of Dementia in the Health and Retirement Study

A subset of HRS (856 adults older than age 70) received in-home clinical assessments of dementia (cognitive, behavioral, and functional status) between August 2001 and December 2003 (approximately 1 year after their HRS evaluation) by a nurse and neuropsychology technician, as part of the ADAMS study. Definitive diagnosis of dementia in the ADAMS sample was assigned by a consensus of clinical experts [11], using international diagnostic criteria for dementia (cognitive or behavioral disorders associated with significant decline in social or occupational functioning) [18]. Hurd and colleagues [10] used HRS data and the clinical diagnosis of dementia from ADAMS to compute predicted probabilities of having dementia 1 year after HRS evaluation based on two ordered probit models [19]. They used variables related to age, education, sex, activities of daily living (ADL), and instrumental activities of daily living (IADL) limitations, cognitive scores from the HRS interview immediately preceding the ADAMS assessment, and changes in ADL and IADL limitations, and in cognitive scores from the two preceding HRS surveys (2 years and 4 years before) in a first model. When the respondents in HRS were not capable to answer questions assessing cognition and behavior, they used items from the Informant Questionnaire on Cognitive Decline in the Elderly [20], consisting of 16 questions that address the respondent's memory and ability to function independently, to compute a second model. Predicted probabilities of dementia from these models were available for 7574 HRS respondents for 2003 [7]. Here, we define high likelihood of dementia in this HRS sample as a predicted probability of dementia greater than .95.

Longitudinal Change in the Created Clusters

Because predicted probabilities of dementia based on the ADAMS clinical diagnosis of dementia are only available in the HRS cohort, we also used longitudinal follow-up in both the HRS and SHARE cohorts to prove the clinical relevance of our unsupervised machine learning classification. To examine longitudinal change in mobility and functional limitations for individuals of the clusters created by our unsupervised machine learning method based on cross-sectional data, we merged data from the classification wave 6 of HRS with data from waves 3, 4, 5, 7, and 8. The longitudinal HRS follow-up covers an average 10-year period, with 6.1 years before classification wave and 3.9 years after. Similarly, data from SHARE classification wave 4 were merged with data from waves 1, 2, 5, and 6 (wave 3 corresponding to SHARE Life study was discarded because of its different protocol). The longitudinal SHARE follow-up covers an average 10.5-year period, with 6.6 years before classification wave and 3.9 years after. Only individuals present in both the earliest and the latest waves were included in both longitudinal datasets (N=10,235 individuals in HRS and N=8245 individuals in SHARE from nine countries including Austria, Belgium, Denmark, France, Germany, Italy, Spain, Sweden, and Switzerland).

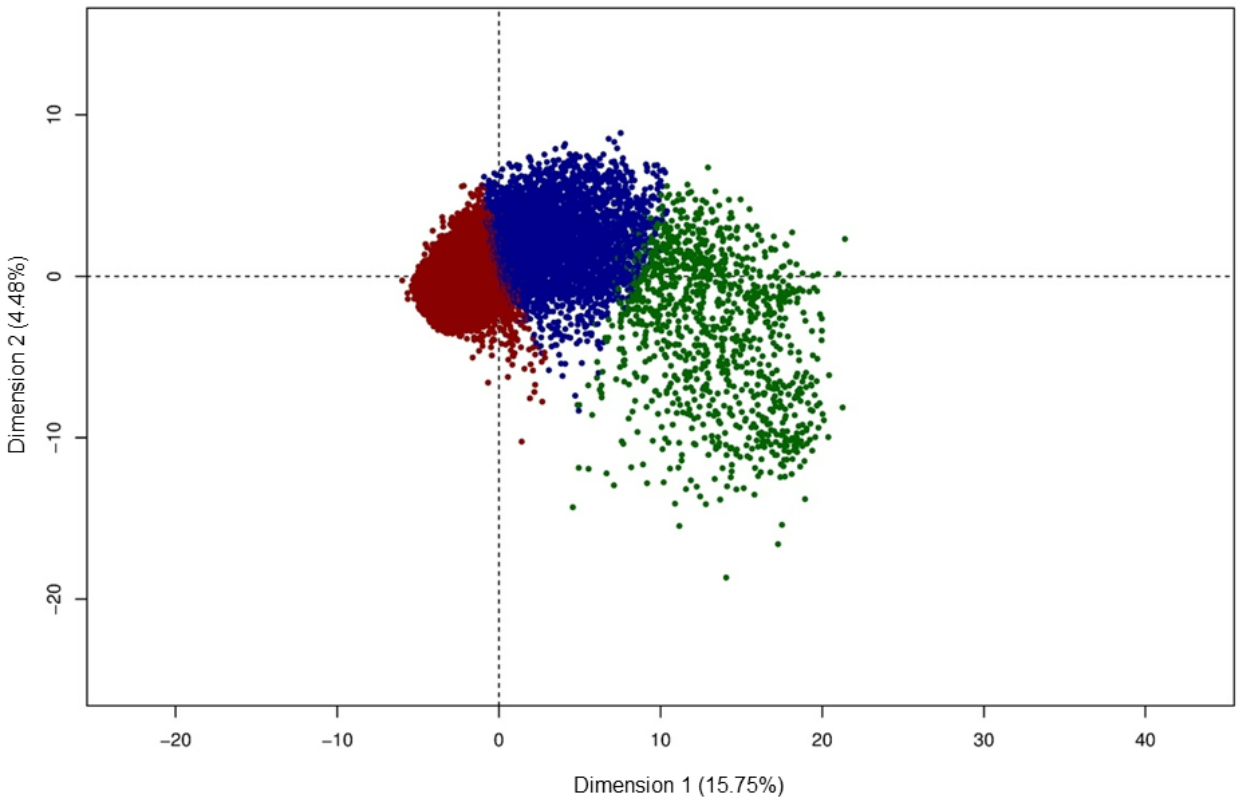
Results

Development Cohort: Health and Retirement Study

We identified three clusters after running unsupervised machine learning classification (Figure 1). Cluster 1 (n=12,231) corresponds to individuals without any functional impairment on both IADL and ADL, and without significant mobility limitation for climbing stairs or walking. Cluster 2 (n=4841) shows moderate mobility limitations, but no functional impairment. Cluster 3 (n=1093) includes individuals with significant functional impairment and mobility limitations. Compared to clusters 1 and 2, individuals in cluster 3 were older, more often women, more often black or Hispanic, less educated, with poorer memory performance, and less likely to be working (Table 1). The percentage of missing values for cognitive variables before imputation differs in the three clusters: 7.30% (893/12,231) in cluster 1, 10.39% (503/4841) in cluster 2, and 59.10% (646/1093) in cluster 3 (P<.001). Similarly, missing values for behavioral measures of depression before imputation amounted to 7.30% (893/12,231) in cluster 1, 10.39% (503/4841) in cluster 2, and 59.01% (645/1093) in cluster 3 (P<.001).

Cluster 3 showed 89.5% sensitivity, 93.3% specificity, 93.1% accuracy overall, and an area under the curve (AUC) of 0.91 for dementia compared to the high likelihood of dementia defined as a predicted probability of dementia >.95 (see Table 2 for other thresholds and for comparison to ADAMS). When cognitive measures were removed from the dataset, classification into cluster 3 showed 88% sensitivity, 93.4% specificity, 93.1% accuracy, and an AUC of 0.91.

Figure 1. Unsupervised hierarchical clustering in the Health and Retirement Study cohort. Scatterplot of the two first dimensions of the principal component analysis (dimension 1 and dimension 2 with explained variance) for individuals in the three clusters (red=cluster 1, blue=cluster 2, green=cluster 3).



When both cognitive and behavioral variables were removed, classification into cluster 3 reached 87% sensitivity, 93.6% specificity, 93.3% accuracy, and an AUC of 0.90. A 98.2% concordant accuracy was found when classifications with and without cognitive variables were compared to each other. The concordant accuracy reached 94.4% between classifications with and without both cognitive and behavioral variables.

Among the 18,165 individuals present at the time of classification (wave 6), 14,670 individuals remained present at wave 8, 1152 individuals were dropouts, and 2343 were deceased. The three clusters differed regarding their survival rate after a 3.9-year period following classification. Survival rate was 94.2% (n=715 deceased) in cluster 1, 80.1% (n=963 deceased) in cluster 2, and 39.2% (n=665 deceased) in cluster 3 (P<.001). Longitudinal change in mobility and functional limitations for the surviving individuals also differed across the three clusters (Figure 2). Individuals in cluster 1 remained stable during the 10-year period of follow-up (from 6 years before classification wave to 4 years after) according to ADL, IADL, and mobility scores. Individuals in cluster 2 showed mild worsening of ADL and IADL scores and moderate worsening of mobility scores during the same 10-year period. Individuals in cluster 3 showed the steepest worsening of ADL, IADL, and mobility scores during the same period.

Validation Cohort: Survey of Health, Ageing and Retirement in Europe

The three clusters, cluster 1 (N=40,223), cluster 2 (N=15,644), and cluster 3 (N=2335), identified in the SHARE dataset (Figure 3) showed similar characteristics compared to those identified for HRS (Table 3). Individuals in cluster 3 were older, more often women, less educated, and less likely to be working compared to clusters 1 and 2. The percentage of missing values for cognitive variables before imputation differed in the three clusters: 1.97% (791/40,223) in cluster 1, 2.85% (446/15,644) in cluster 2, and 18.67% (436/2335) in cluster 3 (P<.001). Missing values for behavioral measures of depression before imputation reached 1.29% (518/40,223) in cluster 1, 1.89% (296/15,644) in cluster 2, and 15.53% (363/2335) in cluster 3 (P<.001). Similar clusters were created if cognitive and behavioral variables were removed from the datasets (96.4% accuracy between classifications with and without cognitive variables; 94.3% accuracy between classifications with and without both cognitive and behavioral variables).

Among the 58,202 individuals present at the time of classification (wave 4), 32,325 were interviewed again at wave 6, whereas 22,406 individuals were dropouts (5861 of those lived in a country not assessed during wave 6) and 3471 were deceased. After a 3.9-year period following classification, the survival rate was 97.1% (n=1067 deceased) in cluster 1, 88.6% (n=1593 deceased) in cluster 2, and 62.2% (n=811 deceased) in cluster 3 (P<.001). The surviving individuals in cluster 1 remained stable during a 10.5-year period (6 years before classification wave to 4 years after), whereas those from cluster 2 showed moderate mobility decline and very mild functional decline. The surviving individuals from cluster 3 showed progressive loss of autonomy according to ADL and IADL scores, and severe mobility impairment over the same period (Figure 2).

Table 1. Demographic and clinical characteristics in the three clusters created by unsupervised machine learning in the Health and Retirement Study (HRS) cohort.

Demographic and clinical characteristics	All (N=18,165)	Cluster 1 (n=12,231)	Cluster 2 (n=4841)	Cluster 3 (n=1093)	*P* value[a]
Age (years), mean (SD)	68.4 (10.5)	66.1 (9.5)	71.4 (10.4)	79.7 (11.3)	<.001
Gender (male), n (%)	7456 (41.05)	5580 (45.62)	1510 (31.19)	366 (33.49)	<.001
Education (years), mean (SD)	12.1 (3.4)	12.8 (3.0)	10.9 (3.5)	10 (4.0)	<.001
Race/ethnicity, n (%)					
White	14,967 (82.39)	10,373 (84.81)	3770 (77.87)	824 (75.39)	<.001
Black	2508 (13.81)	1423 (11.63)	866 (17.89)	219 (20.04)	<.001
Hispanic	1472 (8.10)	886 (7.24)	465 (9.61)	121 (11.07)	<.001
Other race/ethnicity	685 (3.77)	434 (3.55)	201 (4.15)	50 (4.57)	.07
Working full time, n (%)	3773 (20.77)	3470 (28.37)	301 (6.22)	2 (0.18)	<.001
Functional characteristics, mean (SD)					
IADL[b] (0-5)	0.4 (1.0)	0 (0.2)	0.5 (0.8)	3.6 (1.4)	<.001
ADL[c] (0-5)	0.4 (1.0)	0 (0.1)	0.6 (0.9)	3.5 (1.4)	<.001
Mobility[d] (0-5)	1.2 (1.5)	0.4 (0.7)	2.5 (1.4)	3.9 (1.4)	<.001
Total word recall (0-20)	9.4 (4.1)	10.5 (3.4)	8.4 (3.5)	2.2 (4.2)	<.001
CES-D[e] (0-8)	1.6 (2.1)	0.9 (1.3)	3.2 (2.2)	3.4 (3.2)	<.001
Clinical characteristics, n (%)					
Ever had high blood pressure	9167 (50.47)	5265 (43.05)	3183 (65.75)	719 (65.78)	<.001
Ever had diabetes	3029 (16.67)	1456 (11.90)	1286 (26.56)	287 (26.26)	<.001
Ever had cancer	2337 (12.87)	1364 (11.15)	788 (16.28)	185 (16.93)	<.001
Ever had lung disease	1473 (8.11)	499 (4.08)	801 (16.57)	173 (15.83)	<.001
Ever had heart disease	4219 (23.23)	1854 (15.16)	1843 (38.07)	521 (47.67)	<.001
Ever had stroke	1567 (8.63)	469 (3.83)	654 (13.51)	444 (40.62)	<.001
Ever had arthritis	10,231 (56.32)	5501 (44,98)	3903 (80.62)	826 (75.57)	<.001
Ever smoked	10,623 (58.48)	7105 (58.09)	2954 (61,02)	564 (51.60)	<.001
Ever drank alcohol	8103 (44.61)	6573 (53.74)	1410 (29.13)	120 (10,98)	<.001
Body mass index, mean (SD)	27.2 (5.4)	26.9 (4.7)	28.3 (6.5)	25.2 (6.4)	<.001

[a]*P* values are from one-way analysis of variance (ANOVA) or chi-square tests as appropriate.

[b]IADL: instrumental activities of daily living, including any difficulty using a telephone, taking medication, handling money, shopping, and preparing meals.

[c]ADL: activities of daily living, including any difficulty bathing, eating, dressing, walking across a room, and getting in or out of bed.

[d]Mobility: any difficulty for walking several blocks, walking one block, walking across the room, climbing several flights of stairs and climbing one flight of stairs.

[e]CES-D: Center for Epidemiological Study of Depression Scale [13].

Table 2. Classification performance of cluster 3 from unsupervised machine learning in the Health and Retirement Study (HRS) cohort compared to various thresholds of predicted probabilities of dementia from Hurd et al's model and to the Aging, Demographics, and Memory Study (ADAMS) clinical diagnosis of dementia.

Classification performance of cluster 3	Predicted probability of dementia[a] (N=7574)				ADAMS clinical diagnosis of dementia (N=834)
	>.50	>.75	>.90	>.95	
Sensitivity (%)	62.9	77.3	86.7	89.5	59.3
Specificity (%)	96.4	95.1	94.2	93.3	93.0
Accuracy (%)	92.1	93.6	93.7	93.1	81.3

[a]Hurd et al's model [10].

Figure 2. Longitudinal change of instrumental activities of daily living (IADL), activities of daily living (ADL), and mobility scores in both Health and Retirement Study (HRS) and Survey of Health, Ageing and Retirement in Europe (SHARE) cohorts. Linear models with date of assessment at each wave as an independent variable were used to depict the longitudinal change of IADL, ADL, and mobility scores in the three clusters (red=cluster 1, blue=cluster 2, green=cluster 3) in both HRS (left) and SHARE (right) cohorts. A 99% confidence interval (gray color) is drawn for each cluster. The year corresponding to the time of classification is indicated by an arrow.

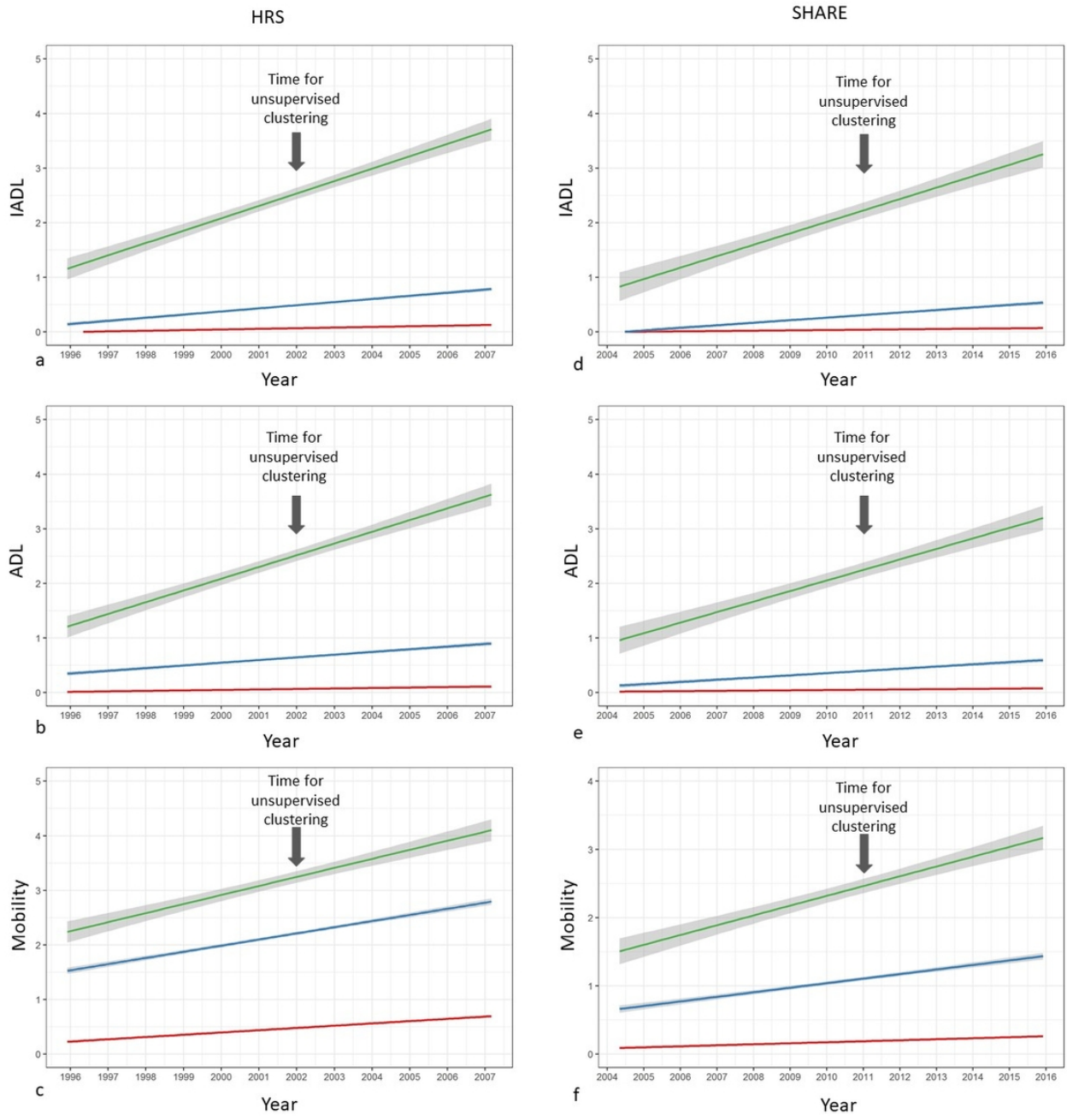


Figure 3. Unsupervised hierarchical clustering in the Survey of Health, Ageing and Retirement in Europe cohort. Scatterplot of the two first dimensions of the principal component analysis (dimension 1 and dimension 2 with explained variance) for individuals in the three clusters (red=cluster 1, blue=cluster 2, green=cluster 3).

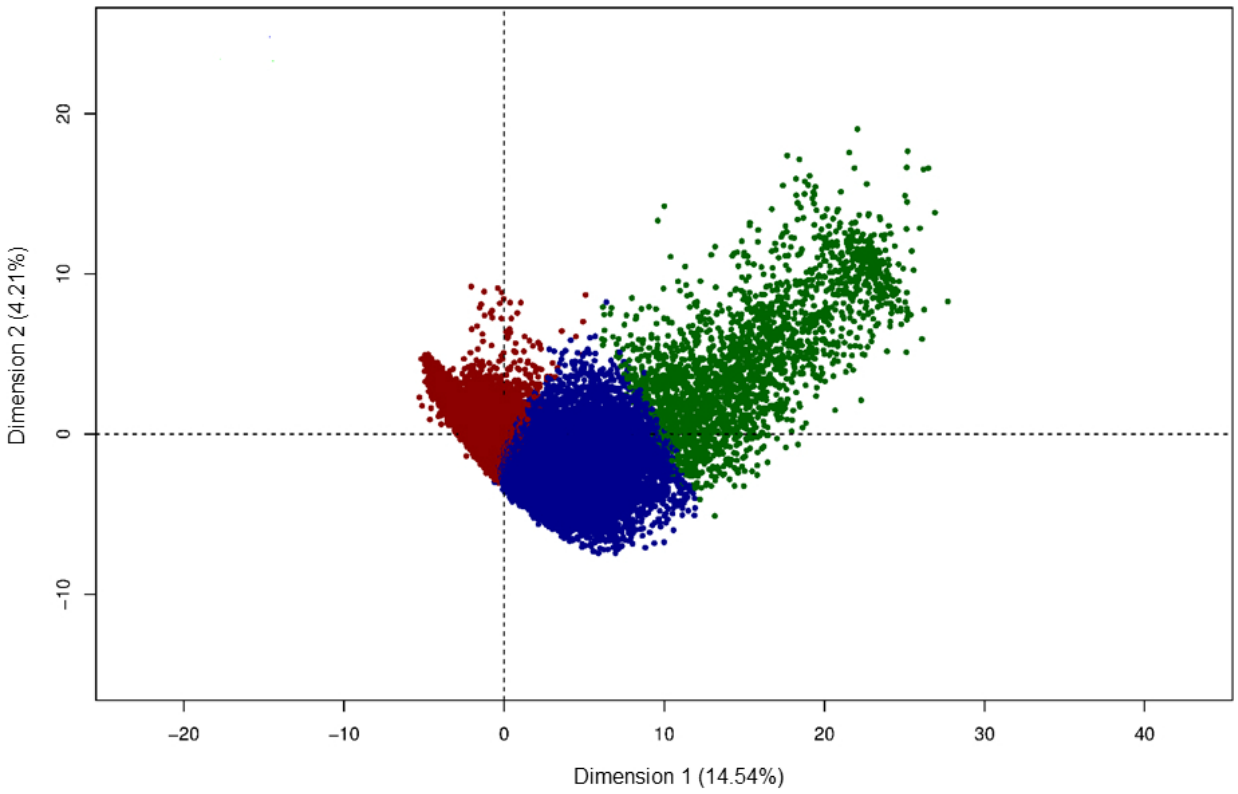


Table 3. Demographic and clinical characteristics in the three clusters created by unsupervised machine learning in the Survey of Health, Ageing and Retirement in Europe (SHARE) cohort.

Demographic and clinical characteristics	All (N=58,202)	Cluster 1 (n=40,223)	Cluster 2 (n=15,644)	Cluster 3 (n=2335)	*P* value[a]
Age (years), mean (SD)	65.4 (10.4)	62.7 (9.0)	70.7 (10.2)	77.4 (10.6)	<.001
Gender (male), n (%)	25,182 (43.26)	19,469 (48.40)	4825 (30.84)	888 (38.03)	<.001
Education (years), mean (SD)	10.3 (4.3)	11.1 (4.1)	8.8 (4.0)	7.8 (4.3)	<.001
Working, n (%)	3889 (6.68)	3591 (8.93)	281 (1.80)	17 (0.73)	<.001
Functional characteristics, mean (SD)					
IADL[b] (0-5)	0.2 (0.8)	0 (0.2)	0.3 (0.6)	3.1 (1.5)	<.001
ADL[c] (0-5)	0.2 (0.8)	0 (0.1)	0.4 (0.7)	3.2 (1.4)	<.001
Mobility[d] (0-4)	0.6 (1.0)	0.1 (0.4)	1.3 (1.0)	3.1 (1.1)	<.001
Total word recall (0-20)	8.9 (3.8)	9.9 (3.4)	7 (3.5)	3.7 (3.8)	<.001
EURO-D[e] (0-12)	2.6 (2.3)	1.8 (1.7)	4.3 (2.4)	5.1 (2.8)	<.001
Clinical characteristics, n (%)					
Ever had high blood pressure	22,848 (39.26)	12,840 (31.92)	8846 (56.55)	1162 (49.76)	<.001
Ever had diabetes	7208 (12.38)	3136 (7.78)	3481 (22.25)	591 (25.31)	<.001
Ever had cancer	3076 (5.29)	1510 (3.75)	1357 (8.67)	209 (8.95)	<.001
Ever had lung disease	3835 (6.59)	1444 (3.59)	2051 (13.11)	340 (14.56)	<.001
Ever had heart disease	7999 (13.74)	2975 (7.40)	4249 (27.16)	775 (33.19)	<.001
Ever had stroke	2547 (4.37)	638 (1.59)	1286 (8.22)	623 (26.68)	<.001
Ever had arthritis	14,192 (24.38)	5797 (14.41)	7347 (46.96)	1035 (44.33)	<.001
Ever smoked	27,097 (46.56)	20,120 (50.02)	6163 (39.40)	814 (34.86)	<.001
Ever drank alcohol	45,893 (78.85)	34,061 (84.68)	10,620 (67.89)	1212 (51.91)	<.001
Body mass index, mean (SD)	26.9 (4.8)	26.4 (4.2)	28.2 (5.9)	26.9 (5.8)	<.001

[a]*P* values are from one-way ANOVAs or chi-square tests as appropriate.

[b]IADL: instrumental activities of daily living, including any difficulty using a telephone, taking medication, handling money, shopping, and preparing meals.

[c]ADL: activities of daily living, including any difficulty bathing, eating, dressing, walking across a room, and getting in or out of bed.

[d]Mobility: any difficulty for walking 100 meters, walking across a room, climbing one flight of stairs, and climbing several flights of stairs.

[e]EURO-D: European Union initiative to compare symptoms of depression scale [14].

Discussion

We used unsupervised machine learning and cross-sectional data from two population-based surveys in the United States and Europe to identify individuals with high likelihood of dementia. Although the clinical diagnosis of dementia usually requires a lengthy and costly process based on human expertise and clinical data, we show that unsupervised machine learning applied to data from population-based surveys provides an accurate estimation of the high probability of dementia, even in the absence of cognitive or behavioral variables. The impact of using unsupervised machine learning in nonmedical datasets would serve to identify older adults with high likelihood of dementia. Being classified into cluster 3 according to our unsupervised machine learning method has clear clinical implications, as shown by the low survival rate during follow-up and the steep functional and mobility declines in the surviving individuals in both the HRS and SHARE cohorts. The higher death rate observed in HRS in comparison to SHARE is likely explained by the older age of the HRS cohort, the better reporting of death date in HRS because of the National Death Index, and the higher number of dropouts in SHARE. Because this unsupervised machine learning method identifies the individuals with worse clinical outcomes, it would be valuable to target those individuals and offer them care including close follow-up or even referral for trials.

Although supervised machine learning is being increasingly used to predict dementia based on clinical data, this study is the first to use unsupervised machine learning and nonclinical data from population-based surveys to identify subjects at risk of dementia. Yet, unsupervised machine learning may be difficult to understand from a clinical perspective. Certain authors compare this purely data-driven method to a “black box” in which the actual mechanisms leading to the outcome remain opaque [21]. In fact, these unsupervised techniques also bring advantages. Because they do not rely on a prespecified clinical outcome (eg, the diagnosis of dementia in a subsample of the

population), they are more flexible than supervised machine learning models and they can be more easily transferred to different types of datasets. Here, this allows classification of individuals from the SHARE cohort where clinical diagnosis of dementia is not available. Moreover, because the unsupervised machine learning algorithm we used is based on PCA, it can assess many variables, such as educational level [22], decline in physical activity [23], slowing gait [24], clinical comorbidities, alcohol consumption, smoking, and weight variations [25], or health care use [26], which are known to be important in the context of dementia. This unsupervised machine learning technique also demonstrates that removing cognitive and behavioral measures from the datasets does not significantly impact the accuracy of the classifications in both HRS and SHARE. The latter result was unexpected given that the current diagnosis criteria of dementia heavily relies on cognitive and behavioral measures [18]. Presumably, this unsupervised machine learning technique is capable of identifying participants with significant decline in social or occupational functioning, often associated with cognitive and behavioral disorders in the context of dementia. It may be of interest for clinicians and researchers because it could allow them to use datasets lacking cognitive or behavioral information such as electronic medical records (EMRs) for studying dementia.

Several aspects of the unsupervised machine learning classification we used may allow for its wide application. In both the HRS and SHARE cohorts, cluster 3 identifies participants who are older; with more cognitive, motor, and functional difficulties; and more likely to show further decline and higher death rate. Thus, this unsupervised machine learning classification technique could be used in other population-based surveys of the HRS family lacking a clinical assessment of dementia such as in ADAMS for the HRS cohort. The longitudinal HRS family studies include the Mexican Health and Aging Study, the English Longitudinal Study of Ageing, the Costa Rican Longevity and Healthy Aging Study, the Korean Longitudinal Study of Aging, the Indonesian Family Life Survey, the Japanese Study of Aging and Retirement, the Asian-African Study on Global AGEing and Adult Health, the Irish Longitudinal Study on Ageing, the Chinese Health and Retirement Longitudinal Study, and the Longitudinal Aging Study in India [7]. Moreover, this unsupervised machine learning algorithm uses cross-sectional data, thus allowing classification of a larger sample of participants at each time point of the survey than the sample that would be constituted if longitudinal data were required. This explains why our method can classify the whole cohort of HRS (N=18,165) compared to the smaller sample (N=7574) in Hurd et al's model [10]. Omitting both cognitive and behavioral variables might further facilitate the inclusion of a larger number of individuals in population surveys. Finally, because it is efficient in two different populations in the United States and Europe even without cognitive or behavioral measures, we expect this classification method to be applicable in other datasets if they constitute representative samples of an aging population.

A possible limitation in this study could be the chosen gold standard to test the accuracy of our classification in HRS cohort. The predicted probabilities of dementia from Hurd et al's models [10] constitute the best reference standard available but they are not definitive. Importantly, Hurd and colleagues provide predicted probabilities of having dementia 1 year after HRS evaluation, whereas our unsupervised machine learning classification directly applies for the time of evaluation, which might account for discrepancies between the two methods. In addition, because Hurd and colleagues used two different models, one when respondents provided answers to cognitive and behavioral measures and another when proxies provided these answers, this might constitute a bias in their predicted probabilities of dementia. The .95 threshold we used for predicted probability of dementia according to Hurd et al's model [10] undoubtedly identifies subjects with high likelihood of dementia, but also misses actual cases of dementia with lower predicted probabilities according to the same model. Indeed, using either lower thresholds of predicted probability of dementia or the actual clinical diagnosis in the smaller sample from ADAMS [11], we obtain similar specificity, but lower sensitivity of cluster 3 regarding the likelihood of dementia (Table 2). Noteworthy, even the diagnosis of dementia from ADAMS and thus the derived Hurd et al's model might suffer from a classification error bias like any clinical assessments [27]. This is why we also use follow-up information related to survival rate and to longitudinal change in functional and mobility scores in the three clusters created after unsupervised machine learning as another way to check for face validity. As expected from patients with dementia, the individuals classified into cluster 3 show a low survival rate and a progressive decline beginning years before the classification time point. Altogether, we acknowledge that this classification method cannot be considered as a diagnosis tool for dementia, or even a dementia-screening instrument, given its moderate sensitivity. Yet, the outcome of this classification, cluster 3, still offers opportunities for new medical applications and new avenues of research in the field of dementia.

Our method could be applied to tackle global health estimates of dementia burden. For example, using the HRS family studies, it could provide a global estimate of dementia across four different continents and an unprecedented cross-country comparison of its socioeconomic consequences, determinants, and risk factors. It could also be applied to other population-based surveys based on different protocols or even to EMRs, often lacking cognitive or behavioral measurements. Whether or not and how the participants at risk of having dementia should be informed after unsupervised machine learning classification raises an ethical issue that would require a large debate. After further validation and using more parsimonious datasets, we expect this unsupervised machine learning classification to impact clinical practice in resource-poor areas with limited primary care access and limited cognitive testing capacities. This technique could support, but not replace, human expertise [28] by identifying groups of individuals with high likelihood of dementia who could then get further clinical assessment and care. Unsupervised machine learning classification applied to existing population datasets or EMRs may help prepare for the global challenge of dementia.

Acknowledgments

LCL, EB, and KY received funding support from the Global Brain Health Institute. KY was supported in part by K24 AG031155. This work was supported by ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL. The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The ADAMS is funded by the National Institute on Aging with the specific aim of conducting a population-based study of dementia. The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857), and FP7 (SHARE-PREP: No 211909, SHARE-LEAP: No 227822, SHARE M4: No 261982). The Gateway to Global Aging Data are funded by the National Institute of Aging (R01 AG030153, RC2 AG036691, R03 AG043052) and developed and maintained by the Program on Global Aging, Health, and Policy, the USC Dornsife Center for Economic and Social Research. LCL had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The sponsors were not involved in designing the study, collecting data, analyzing data, reviewing, or approving the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Variables used in the Health and Retirement Study (HRS) and Survey of Health, Ageing and Retirement in Europe (SHARE) datasets. Asterisk indicates variables related to the past year in SHARE, to the 2 past years in HRS; § to those omitted in datasets without cognitive variables; and # to those omitted in the datasets without behavioral variables.

[PDF File (Adobe PDF File), 57KB - jmir_v20i7e10493_app1.pdf]

References

1. Norton S, Matthews FE, Brayne C. A commentary on studies presenting projections of the future prevalence of dementia. BMC Public Health 2013 Jan 02;13:1 [FREE Full text] [doi: 10.1186/1471-2458-13-1] [Medline: 23280303]
2. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging Ment Health 2011 Nov;15(8):978-984. [doi: 10.1080/13607863.2011.596805] [Medline: 21777080]
3. Dias A, Patel V. Closing the treatment gap for dementia in India. Indian J Psychiatry 2009 Jan;51 Suppl 1:S93-S97 [FREE Full text] [Medline: 21416026]
4. Nakamura AE, Opaleye D, Tani G, Ferri CP. Dementia underdiagnosis in Brazil. Lancet 2015 Jan 31;385(9966):418-419. [doi: 10.1016/S0140-6736(15)60153-2] [Medline: 25706975]
5. Jedenius E, Wimo A, Strömqvist J, Jönsson L, Andreasen N. The cost of diagnosing dementia in a community setting. Int J Geriatr Psychiatry 2010 May;25(5):476-482. [doi: 10.1002/gps.2365] [Medline: 19685441]
6. World Health Organization. Dementia: A Public Health Priority. Geneva: World Health Organization; 2012. URL: http://apps.who.int/iris/bitstream/handle/10665/75263/9789241564458_eng.pdf;jsessionid=8746C61E06238D457433FB5DCB73353C?sequence=1 [accessed 2018-07-02] [WebCite Cache ID 70cL4R7Jh]
7. Gateway to Global Aging Data. URL: https://g2aging.org/ [accessed 2018-03-23] [WebCite Cache ID 6y8PmdVVs]
8. Obermeyer Z, Emanuel EJ. Predicting the future-big data, machine learning, and clinical medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
9. Health and Retirement Study. URL: http://hrsonline.isr.umich.edu/ [accessed 2018-03-23] [WebCite Cache ID 6y8abpMa1]
10. Hurd MD, Martorell P, Delavande A, Mullen KJ, Langa KM. Monetary costs of dementia in the United States. N Engl J Med 2013 Apr 4;368(14):1326-1334 [FREE Full text] [doi: 10.1056/NEJMsa1204629] [Medline: 23550670]
11. Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Ofstedal MB, et al. The Aging, Demographics, and Memory Study: study design and methods. Neuroepidemiology 2005;25(4):181-191. [doi: 10.1159/000087448] [Medline: 16103729]
12. SHARE: Survey of Health, Ageing and Retirement in Europe. URL: http://www.share-project.org/ [accessed 2018-03-23] [WebCite Cache ID 6y8apxMp7]
13. CESD-R: Center for Epidemiologic Studies Depression Scale Revised. URL: http://cesd-r.com/ [accessed 2018-03-23] [WebCite Cache ID 6y8ahBgWy]
14. Prince MJ, Reischies F, Beekman AT, Fuhrer R, Jonker C, Kivela SL, et al. Development of the EURO-D scale--a European, Union initiative to compare symptoms of depression in 14 European centres. Br J Psychiatry 1999 Apr;174:330-338. [Medline: 10533552]
15. Josse J, Husson F. missMDA: a package for handling missing values in multivariate data analysis. J Stat Soft 2016;70(1):1-31. [doi: 10.18637/jss.v070.i01]

16. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Soft 2008;25(1):1-18. [doi: 10.18637/jss.v025.i01]
17. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Soft 2014;61(6):1-36. [doi: 10.18637/jss.v061.i06]
18. Blazer D. Neurocognitive disorders in DSM-5. Am J Psychiatry 2013 Jun;170(6):585-587 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13020179] [Medline: 23732964]
19. Greene W. Econometric Analysis (8th Edition). London: Pearson; 2018.
20. Jorm AF. A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. Psychol Med 1994 Feb;24(01):145-153. [doi: 10.1017/S003329170002691X]
21. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA 2017 Aug 08;318(6):517-518. [doi: 10.1001/jama.2017.7797] [Medline: 28727867]
22. Xu W, Tan L, Wang H, Tan M, Tan L, Li J, et al. Education and risk of dementia: dose-response meta-analysis of prospective cohort studies. Mol Neurobiol 2016 Dec;53(5):3113-3123. [doi: 10.1007/s12035-015-9211-5] [Medline: 25983035]
23. Sabia S, Dugravot A, Dartigues J, Abell J, Elbaz A, Kivimäki M, et al. Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. BMJ 2017 Jun 22;357:j2709 [FREE Full text] [Medline: 28642251]
24. Rosso AL, Verghese J, Metti AL, Boudreau RM, Aizenstein HJ, Kritchevsky S, et al. Slowing gait and risk for cognitive impairment: the hippocampus as a shared neural substrate. Neurology 2017 Jul 25;89(4):336-342. [doi: 10.1212/WNL.0000000000004153] [Medline: 28659421]
25. Vos SJ, van Boxtel MP, Schiepers OJG, Deckers K, de Vugt V, Carrière I, et al. Modifiable risk factors for prevention of dementia in midlife, late life and the oldest-old: validation of the LIBRA Index. J Alzheimers Dis 2017;58(2):537-547. [doi: 10.3233/JAD-161208] [Medline: 28453475]
26. LaMantia MA, Stump TE, Messina FC, Miller DK, Callahan CM. Emergency department use among older adults with dementia. Alzheimer Dis Assoc Disord 2016;30(1):35-40 [FREE Full text] [doi: 10.1097/WAD.0000000000000118] [Medline: 26523710]
27. Plassman BL, Khachaturian AS, Townsend JJ, Ball MJ, Steffens DC, Leslie CE, et al. Comparison of clinical and neuropathologic diagnoses of Alzheimer's disease in 3 epidemiologic samples. Alzheimers Dement 2006 Jan;2(1):2-11. [doi: 10.1016/j.jalz.2005.11.001] [Medline: 19595851]
28. Chen JH, Asch SM. Machine learning and prediction in medicine-beyond the peak of inflated expectations. N Engl J Med 2017 Jun 29;376(26):2507-2509. [doi: 10.1056/NEJMp1702071] [Medline: 28657867]

Abbreviations

ADAMS: Aging, Demographics, and Memory Study
ADL: activities of daily living
ANOVA: analysis of variance
AUC: area under the curve
CES-D: Center for Epidemiological Study of Depression Scale
EMR: electronic medical record
EURO-D: European Union initiative to compare symptoms of depression scale
HRS: Health and Retirement Study
IADL: instrumental activities of daily living
PCA: principal component analysis
SHARE: Survey of Health, Ageing and Retirement in Europe

Edited by G Eysenbach; submitted 25.03.18; peer-reviewed by D Trepel, M Pino; comments to author 25.04.18; revised version received 08.05.18; accepted 08.05.18; published 09.07.18.

Please cite as:
Cleret de Langavant L, Bayen E, Yaffe K
Unsupervised Machine Learning to Identify High Likelihood of Dementia in Population-Based Surveys: Development and Validation Study
J Med Internet Res 2018;20(7):e10493
URL: http://www.jmir.org/2018/7/e10493/
doi:10.2196/10493
PMID:29986849

Original Paper

A Decade of Veteran Voices: Examining Patient Portal Enhancements Through the Lens of User-Centered Design

Kim M Nazi[1*], PhD; Carolyn L Turvey[2,3*], PhD; Dawn M Klein[2,3*], MSW; Timothy P Hogan[4,5*], PhD

[1]Veterans and Consumers Health Informatics Office, Office of Connected Care, Veterans Health Administration, US Department of Veterans Affairs, Coxsackie, NY, United States

[2]Iowa City VA Health Care System, Comprehensive Access and Delivery Research and Evaluation Center, Iowa City, IA, United States

[3]Department of Psychiatry, The University of Iowa Carver College of Medicine, Iowa City, IA, United States

[4]Center for Healthcare Organization and Implementation Research (CHOIR), Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

[5]Division of Health Informatics and Implementation Science, Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[*]all authors contributed equally

Corresponding Author:
Kim M Nazi, PhD
Veterans and Consumers Health Informatics Office, Office of Connected Care
Veterans Health Administration
US Department of Veterans Affairs
44 Mountain View Drive
Coxsackie, NY, 12051
United States
Phone: 1 518 703 6330
Email: kmnazi@gmail.com

Abstract

Background: Health care systems have entered a new era focused on patient engagement. Patient portals linked to electronic health records are recognized as a promising multifaceted tool to help achieve patient engagement goals. Achieving significant growth in adoption and use requires agile evaluation methods to complement periodic formal research efforts.

Objective: This paper describes one of the implementation strategies that the Department of Veterans Affairs (VA) has used to foster the adoption and sustained use of its patient portal, My Health*e*Vet, over the last decade: an ongoing focus on user-centered design (UCD). This strategy entails understanding the users and their tasks and goals and optimizing portal design and functionality accordingly. Using a case study approach, we present a comparison of early user demographics and preferences with more recent data and several examples to illustrate how a UCD can serve as an effective implementation strategy for a patient portal within a large integrated health care system.

Methods: VA has employed a customer experience analytics (CXA) survey on its patient portal since 2007 to enable ongoing direct user feedback. In a continuous cycle, a random sample of site visitors is invited to participate in the Web-based survey. CXA model questions are used to track and trend satisfaction, while custom questions collect data about users' characteristics, needs, and preferences. In this case study, we performed analyses of descriptive statistics comparing user characteristics and preferences from FY2008 (wherein "FY" means "fiscal year") to FY2017 and user trends regarding satisfaction with and utilization of specific portal functions over the last decade, as well as qualitative content analysis of user's open-ended survey comments.

Results: User feedback has guided the development of enhancements to core components of the My Health*e*Vet portal including available features, content, interface design, prospective functional design, and related policies. Ten-year data regarding user characteristics and portal utilization demonstrate trends toward greater patient engagement and satisfaction. Administration of a continuous voluntary Web-based survey is an efficient and effective way to capture veterans' voices about who they are, how they use the patient portal, needed system improvements, and desired additional services.

Conclusions: Leveraging "voice-of-the-customer" techniques as part of patient portal implementation can ensure that such systems meet users' needs in ways that are agile and most effective. Through this strategy, VA has fostered significant adoption and use of My Health*e*Vet to engage patients in managing their health.

(J Med Internet Res 2018;20(7):e10413) doi:10.2196/10413

KEYWORDS

patient portal; user-centered design; eHealth; veteran

Introduction

Background

Health care systems have entered a new era focused on patient engagement [1-3] described by enthusiasts as "the holy grail of health care" [4] and the "blockbuster drug of the century" [5]. Patient engagement strategies are designed to empower patients to play a more active role in their health care and make informed decisions, improve the patient experience, increase patient satisfaction, and achieve better health outcomes. Patient portals linked to electronic health records (EHRs) are recognized as a promising multifaceted tool to help achieve these patient engagement goals [6-9]. However, the adoption and sustained use of portals has generally fallen short of initial optimism [10-13] even in light of the significant growth in EHRs and tethered patient portals incentivized by Meaningful Use [14]. Positive benefits of portal use have been demonstrated [15-19], and the OpenNotes movement [20] has promoted patient engagement through health records transparency by enabling patient access to provider notes. Evidence indicates that such access improves communication and trust, patient safety, and, potentially, patient outcomes [21-25]. Two large integrated health care systems that launched tethered patient portals in 2003 with significant patient adoption and sustained use are Kaiser Permanente (KP) and the Department of Veterans Affairs (VA). KP's portal, My Health Manager, is used by more than 5 million members, representing about 70% of adult KP members [26]. VA's patient portal, My Health*e*Vet, has more than 4 million registered users (69% of VA patients receiving health care services in FY 2017 [wherein "FY" means "fiscal year"]), with 2.5 million authenticated Premium accounts (42% of VA patients receiving health care services) required for access to all portal features [27]. To better understand what elements are driving this adoption and sustained use, an implementation case study approach is warranted. One of the implementation strategies that are critical to foster the adoption and sustained use of patient portals is an ongoing focus on user-centered design (UCD). This is often accomplished as part of periodic research studies; however, more timely and agile methods are needed to design and evaluate patient portals.

User-Centered Design

UCD is a design philosophy and evaluation process that focuses on the end user's characteristics, needs, preferences, and limitations throughout the design process and development lifecycle [28]. The emphasis of UCD is on understanding the end users and their tasks and goals and optimizing the product to enable the users to fulfill these, rather than requiring users to adapt to the designer's preferences [29]. UCD of eHealth applications, such as patient portals, necessitates ongoing assessment of user characteristics and preferences and incorporation of assessment insights into ongoing portal development and enhancements. This process includes focusing on what features are considered to be most essential by users [30]. Published compilations of implementation strategies have called for the further development of processes like UCD as a means to obtain and use patient or consumer feedback to support the adoption of innovations and practice change efforts in health care and other settings [31].

VA has used various methods over the last decade to achieve UCD for My Health*e*Vet; among them, the principal method has been a continuous, voluntary, and anonymous survey of end users. As a complement to periodic formal research studies [32,33], this ongoing assessment offers the advantage of rapid continuous feedback, which is part of a cyclical process for improvement that entails understanding users, eliciting their input, identifying changes or future design implications, deploying enhancements, and then obtaining feedback to evaluate these enhancements. This method enables VA to obtain ongoing direct feedback from veterans, which can then be leveraged to improve the patient experience.

About the Department of Veterans Affairs' Patient Portal

VA is the largest integrated health care system in the United States and has been a pioneer in enabling patients to access and download their VA medical record data using the Blue Button feature [34,35]. This includes OpenNotes, which are known in My Health*e*Vet as VA Notes and contain both clinical and mental health providers' notes [36]. The My Health*e*Vet patient portal [37] is tethered to the VA EHR and provides a suite of Web-based tools. Veterans self-register to create a basic account and can then self-enter information into their personal health record and access health education resources. VA patients who are matched by the system via the Master Veteran Index are automatically upgraded to an Advanced account and can request VA prescription refills. Patients who complete a one-time process of identity authentication (in person or Web-based) are upgraded to a Premium account and can then access all portal features, including access to health record information and Secure Messaging with VA health care professionals.

Use of My Health*e*Vet continues to grow. In fiscal year (FY)2017, portal user activity demonstrated significant increases compared with that in FY2016, including a 20.7% increase in Web-based prescription refills, a 33.9% increase in Secure Messaging exchanges between VA patients and their health care team, and a 38.7% increase in use of the VA Blue Button feature [27].

In this paper, we examine one of the implementation strategies that VA has used to foster adoption and sustained use of its patient portal over the last decade: an ongoing focus on UCD. This includes iterative use of survey and operational data with user interface redesign to meet the needs and preferences of veteran users. We describe the organization's implementation strategy for agile UCD and present unique 10-year data on user adoption, characteristics, and utilization to demonstrate trends toward greater patient engagement and satisfaction. Following

an initial analysis of portal users and their preferences in 2007 [38], we compared the characteristics of patient portal users one decade later and used a case study approach to present several examples of how user preferences and continuous feedback have informed the evolution of VA's patient portal.

Methods

Since 2007, VA has used the ForeSee customer experience analytics (CXA) survey tool for the direct measurement of customer satisfaction and prioritization of enhancements. The CXA survey is a standardized method of measuring and monitoring customer satisfaction based on the American Customer Satisfaction Index [39]. The survey methodology uses a psychometric "voice-of-the-customer" technique to assess consumer drivers of satisfaction (look and feel, navigation, site information, site performance, and task processes) and prioritize areas of improvement. In the CXA model, scores are based on data from randomized voluntary Web-based surveys and are reported on a scale of 0 to 100, indicating less to more customer satisfaction. Multiple item measures are combined algorithmically to compile a satisfaction index each time an adequate quantity of data has been collected through completed surveys [40]. The survey tool for the My Health*e*Vet portal includes standard questions, to allow for trend analysis of core components such as overall satisfaction, and user experience of navigation. The inclusion of custom questions on an as-needed basis further enables the collection of rich data about user demographics, needs, and preferences to address specific and time-sensitive evaluation topics and to inform ongoing design and development efforts.

The CXA survey is conducted with all veterans using My Health*e*Vet and is, therefore, a nationwide sample of veteran My Health*e*Vet users. The survey is implemented on the My Health*e*Vet portal as a Web-based pop-up browser window inviting a random sample of site visitors to participate. A persistent cookie prevents site visitors who received the survey invitation from being invited again for 90 days. When visitors accept the invitation, the survey presents when they leave the site. The loyalty factor, currently 4 pages, ensures that respondents have experienced multiple pages on the site before being prompted to participate in the survey. The sampling percentage, set at 13% in FY2008 and later changed to 4% in FY2010 due to the large amount of data being collected and increasing survey completion rates, ensures that a minimum number of site visitors are surveyed in order to reduce respondent burden while enabling the collection of adequate data.

This paper presents selected analyses of the CXA survey data collected over a course of 10 years, including a comparison of data collected early in the implementation of My Health*e*Vet (FY2008) to more recent data (FY2017), to examine the characteristics of patient portal users and their preferences.

Data analysis is primarily descriptive and based on forced-choice responses. Analysis of open-ended comments includes a combination of traditional qualitative techniques [41,42] along with keyword clustering to group related comments for further analysis. A variety of strategies are used to then translate insights into iterative improvements, including ongoing data reviews, requirement elaboration, design sessions with key stakeholders, and review of user feedback after deployment of enhancements.

Results

Overview

We first present a recent summary of user demographics and characteristics and patterns of portal use and relevant comparisons to previous data. Following our case study approach, we then provide selected examples from the My Health*e*Vet evaluation program to illustrate how different assessments that capture the voice of the customer have directly informed the evolution of the portal and the addition of new functionality. For FY2008 (October 1, 2007-September 30, 2008), of the surveys presented to site visitors, 17.1% (100,069/585,039) were completed. For FY2017 (October 1, 2016-September 30, 2017), of the surveys presented to site visitors, 68.9% (100,555/146,023) were completed. As completion rates increased over the last decade, the sampling rate was reduced in FY2010 from 13% to 4% in order to minimize respondent burden.

User Demographics and Characteristics

Table 1 provides a comparison of user demographics and characteristics for all survey respondents in FY2017 and FY2008. In FY2017, 97% (97,538/100,555) of respondents were veterans compared with 93% (93,064/100,069) in FY2008. Respondents reported having completed higher levels of education, with 40% (39,990/99,974) being college graduates, completing some postgraduate school, or having a graduate or professional degree in FY2017 compared with 34% (732/2152) in FY2008. The proportion of male respondents increased slightly to 93% (90,507/97,319) in FY2017. In FY2017, respondents were generally older, with 64% (59,819/93,467) in the age range of 60-74 years compared with 47% (14,563/30,984) in FY2008; furthermore, 17% (15,889/93,467) of respondents in FY2017 were older than 75 years. This shift in age is also shown in Figure 1.

While 60% (60,042/100,069) of users in FY2008 reported their period of military service as the Vietnam War, this increased to 67% (67,372/100,555) of users in FY2017. Fewer users self-reported their internet ability as advanced in FY2017 (32,235/53,725, 60%) than in FY2008 (37,848/55,658, 68%), whereas more users reported it as intermediate in FY2017 (19,341/53,725, 36%) than in FY2008 (16,141/55,658, 29%). A greater proportion of respondents reported better health in FY2017, with 34% (33,323/98,007) reporting fair or poor health in FY2017 compared with 39% (15,723/40,315) in FY2008. Although the FY2008 survey did not ask users about health conditions, responses in FY2017 revealed a high prevalence of chronic conditions, including high blood pressure (15,045/22,795, 66%), high cholesterol (14,133/22,795, 62%), arthritis (13,677/22,795, 60%), chronic pain (10,714/22,795, 47%), diabetes (8434/22,795, 37%), stomach or gastrointestinal problems (8434/22,795, 37%), and heart problems (8434/22,795, 37%).

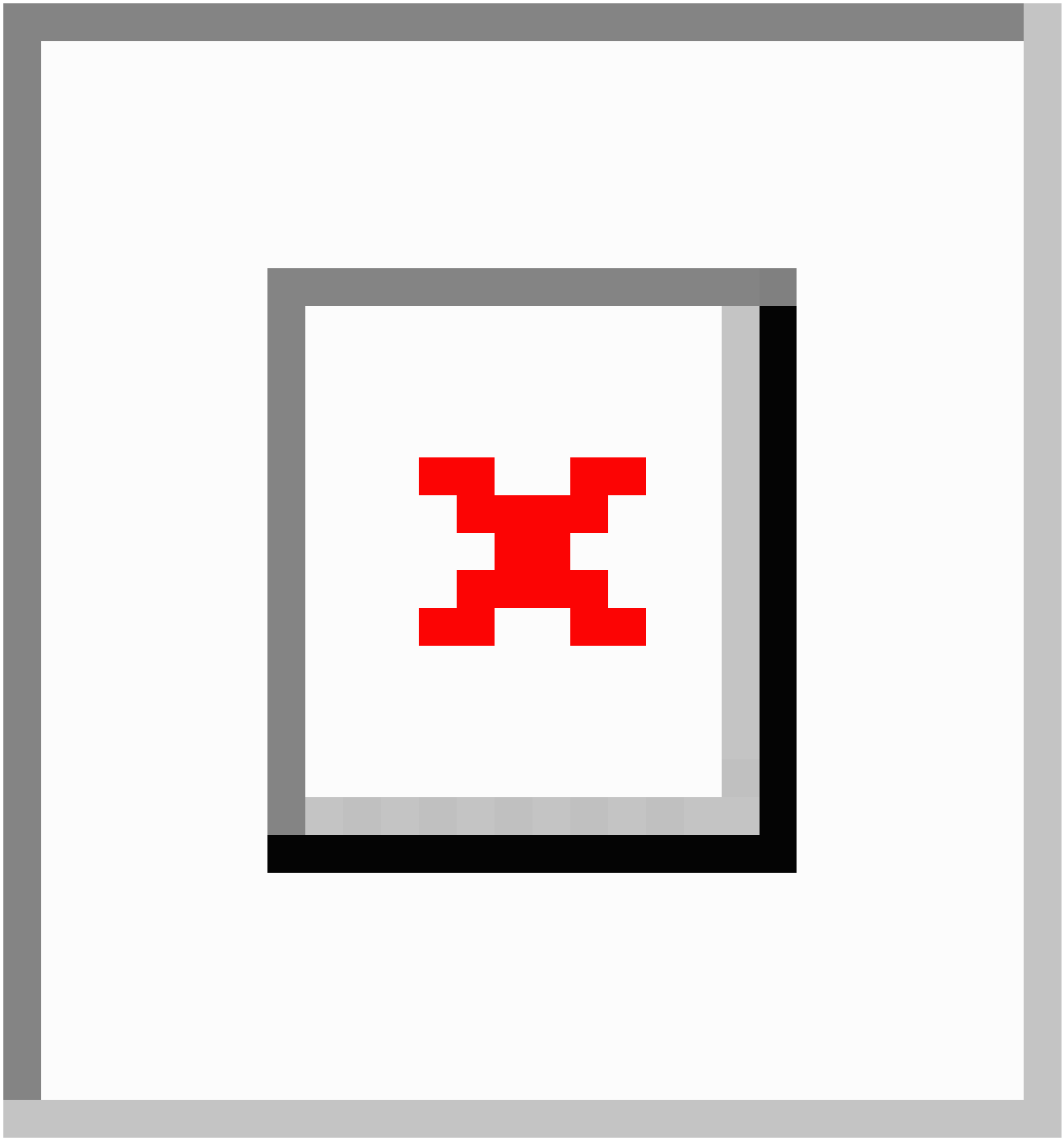
Table 1. Demographics and characteristics. FY: fiscal year. N/A: Not applicable. VA: Department of Veterans Affairs.

Type	FY2017	FY2008
Role[a], n (%)	**100,555**	**100,069**
Veteran	97	93
Family member	3	5
Veteran Service Organization	1	1
National Guard or Reserve	1	N/A
General public	1	<1
Other role	1	1
VA employee	1	1
Non-VA federal employee	<1	1
Caregiver (other than family)	<1	N/A
State or local government	<1	N/A
Active duty	<1	<1
News Media	N/A	<1
Highest level of education, n (%)	**99,974**	**2154**
Did not complete high school	3	2
High school graduate	13	17
Some college or vocational school	44	44
College graduate	21	19
Some postgraduate school	6	5
Graduate or professional degree	13	10
Health conditions[a], n (%)	**22,795**	**N/A**
High blood pressure	66	N/A
High cholesterol	62	N/A
Arthritis of any kind	60	N/A
Chronic pain	47	N/A
Diabetes	37	N/A
Stomach or gastrointestinal problems	37	N/A
Heart problems	37	N/A
Mental health or psychiatric condition	34	N/A
Cancer of any kind	29	N/A
Lung problems (including asthma)	25	N/A
Neurological disorders	13	N/A
Other	12	N/A
Prefer not to answer	2	N/A
Age, n (%)	**93,467**	**30,984**
Under 20	<1	<1
20-24	<1	<1
25-29	<1	<1
30-34	<1	1
35-39	1	2
40-44	1	4

Type	FY2017	FY2008
45-49	3	6
50-54	5	10
55-59	8	18
60-64	13	26
65-69	28	14
70-74	23	7
75-79	9	5
80-84	5	3
85 or older	3	1
Gender, n (%)	**97,319**	**31,020**
Male	93	91
Female	7	9
Self-reported health status, n (%)	**98,007**	**40,315**
Excellent	4	5
Very good	21	18
Good	41	38
Fair	27	29
Poor	7	10
Self-reported internet ability, n (%)	**53,725**	**55,658**
Beginner	4	4
Intermediate	36	29
Advanced	60	68

[a]Multiple categories may be selected.

Figure 1. User age ranges for fiscal year (FY)2008 and FY2017.



Portal Access Patterns and Usage

Table 2 provides a summary of survey respondents' self-reported portal access patterns and usage. While the proportion of survey respondents who use VA health care services remained the same from FY2008 to FY2017 (96%), in FY2017 50% (47,066/94,132) of respondents stated that they also use a community non-VA provider. When asked about travel time to the nearest VA facility, 32% respondents reported it to be less than 30 minutes in both FY2008 and FY2017; however, a greater proportion noted fewer minutes of travel time in FY2017 than in FY2008. For example, 37% (19,902/53,788) respondents in FY2008 reported a travel time greater than 60 minutes to the nearest VA facility compared with 24% (12,075/50,313) in FY2017. The proportion of respondents who have a Premium account, offering them access to all portal services, increased significantly from 60% (56,884/94,806) in FY2008 to 77% (73,001/94,806) in FY2017. While a greater number of respondents were first time users in FY2008 (12,074/100,617, 12%) than in FY2017 (4022/100,555, 4%), respondents reported using the portal more frequently in FY2017, with 46% (45,255/100,555) using it about once a month and 29% (29,161/100,555) using it about once a week. When asked about the length of use in FY2017, 63% (63,349/100,5455) respondents reported having used My Health*e*Vet for more than 2 years.

User Preferences and Responsive Design

In keeping with the UCD process, VA has used direct veteran feedback about preferences obtained via the CXA survey to shape the identification and prioritization of portal improvements. In this section, we describe how different types of user feedback have directly informed enhancements to the core components of the system including available features, interface design, content, policy, and prospective functional design of a new feature.

Additional Services Desired

UCD principles focus on identifying what features users consider to be essential. One survey question that has been crucial in getting feedback to prioritize portal enhancements over the last decade has been "What additional services would you like to see on My Health*e*Vet?" As shown in Table 3, additional services desired by users in FY2008 included the ability to view (79,892/92,160, 87%) or schedule (68,395/92,160, 74%) VA Appointments, access information from the VA medical record (67,714/92,160, 73%), and Web-based secure communication with my doctor (58,878/92,160, 64%). Each of these features was subsequently added to the portal (Table 4).

Secure Messaging implementation began in 2008, which enabled secure Web-based communication with VA health care teams, with the full national release to all VA primary care providers in 2012. Veterans could then also use Secure Messaging to request VA Appointments. The ability to view upcoming VA Appointments was deployed in 2011, with appointment email reminders added in 2015. Building on early access to VA Medication History, VA incrementally expanded the types of information from the VA medical record available in My Health*e*Vet, for example, preventative Wellness Reminders (2009), VA Chemistry or Hematology Lab Results (2011), VA Immunizations (2012), VA Notes including mental health notes (2013), a more comprehensive Medication List that includes patient-reported non-VA medications (2016), Surgical and Clinical Procedure Notes (2017), and VA Medical Images and Reports (2017).

Additional services desired by users in FY2017 included the ability to schedule or change VA Appointments directly (52%), a list of health care providers and their contact information (44%), a tool to determine whether different medications are safe when taken together (26%), and the ability to view and pay VA bills or copayments (25%). The ability to schedule or change VA Appointments directly was piloted in FY2017 and is being rolled out to all VA facilities in FY2018. The enhanced VA Health Summary (2017) provides VA patients with a list of their primary health care providers, which will be expanded in FY2018 to include their contact information. Although VA has not yet invested in the development of tools to check medications for potential interactions; this enhancement is being given further consideration in FY2018. In addition, the ability to view a VA Patient Statement and remit payment is also being developed and scheduled for pilot testing in FY2018.

Patient-Identified Main Improvements

In addition to eliciting user feedback on additional services desired, the CXA survey also invites open-ended comments in response to the question: "What is the main improvement that you would suggest for the My Health*e*Vet website?" Below we offer examples of how these comments have led to user-directed improvements.

With the expansion of Lab and Test Results and the addition of VA Notes in January 2013, one theme that surfaced in the ensuing months was veterans' desire for more timely access to this information. These comments were crucial in driving VA policy change to reduce the hold period for lab results and progress notes from 7 calendar days after verification to 3 calendar days. This policy change was implemented in June 2013.

To complement the prioritization of known desired additional services by users, open-ended main improvement comments also allow veterans to suggest needed functional enhancements in their own words. In October 2013, thematic analysis of free-text comments identified the need for multiple functional enhancements including the ability to track delivery of the filled prescriptions, the desire to be notified before automatic log out when the user session was nearing time-out, and the need for improved navigation to complete common tasks. The ability to track delivery of mailed prescriptions by opting-in to receive an email notification was deployed in 2015. Other functional enhancements (session time-out warning, improved navigation, and reduced number of steps to complete common tasks) became core requirements for a major website redesign project. The session time-out warning and ability to extend the session time was deployed as VA migrated to a content management system in October 2016. The incremental deployment of website redesign in October 2016 and September 2017 was significantly informed by veteran main improvement comments:

> *Publishing labs and notes within 24 hours of a lab or health visit. Waiting a week for lab results, or a week for Dr and nurse notes is absurd, given that the health problem is "right now," not right now + seven days, especially when Dr's notes are also instructions for post visit procedures, such as when and how much meds to take, or "If it hasn't improved in three days" see me. Not everyone is "present" at the end of a visit due mostly to anxieties surrounding the visit.*

Table 2. Access patterns. FY: fiscal year. N/A: not applicable. VA: Department of Veterans Affairs.

Respondent Characteristic	FY2017	FY2008
Use VA health care services, n (%)	**98,007**	**29,528**
Yes	96	96
No	3	4
Not sure	1	N/A
Use community non-VA providers, n (%)	**94,132**	**N/A**
Yes	50	N/A
No	47	N/A
Not sure	3	N/A
Premium My HealtheVet account[a], n (%)	**94,806**	**100,617**
Yes	77	60
No	9	24
Not sure	15	15
Not applicable	N/A	1
Travel time to nearest VA facility[a], n (%)	**50,313**	**53,788**
Less than 30 min	32	32
30-60 min	43	32
61-90 min	14	20
91 min to 2 h	6	9
Over 2 h	4	8
Not sure	N/A	1
Frequency of use, n (%)	**100,555**	**100,617**
Daily or more than once a day	5	5
About once a week	29	25
About once a month	46	49
About every 6 mo	9	5
Less than every 6 mo	4	3
First time	4	12
Not sure or Do not recall	2	N/A
Length of use, n (%)	**100,555**	**N/A**
Less than 6 mo	9	N/A
6 mo-less than 1 y	6	N/A
1-2 y	19	N/A
More than 2 y	63	N/A
Not sure or Do not recall	3	N/A

[a]Percentages do not add to 100 due to rounding.

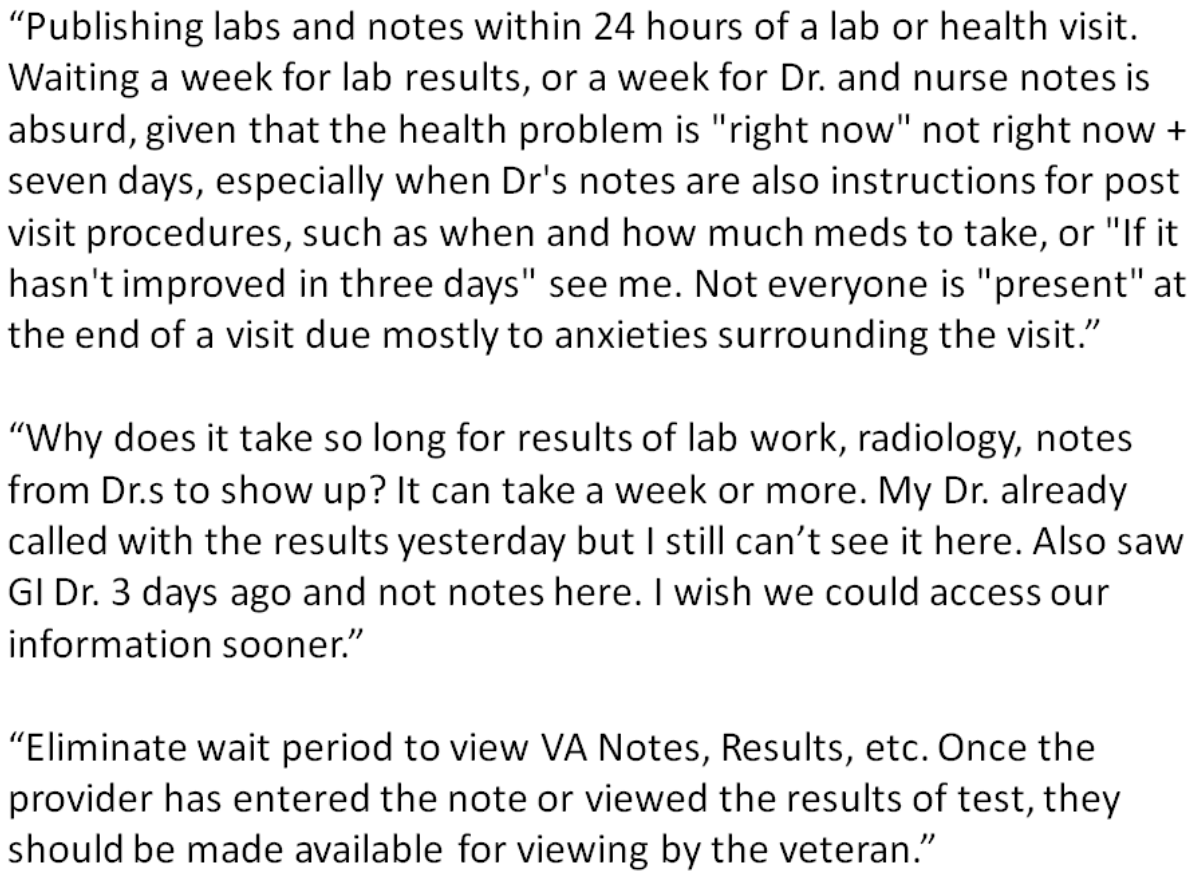
Table 3. Additional services desired. FY: fiscal year. VA: Department of Veterans Affairs.

Service	n (%)
FY2017	**88,308**
Schedule or change my VA appointments	45,695 (52)
View a list of my VA health care providers and their contact information	38,489 (44)
Check to determine if my different medications are safe taken together	22,710 (26)
View or pay my VA bills or copayments	21,768 (25)
Use a mobile app for My Health*e*Vet	13,823 (16)
Advance check-in for my VA clinic visits	12,677 (14)
Authorize sharing information with my Non-VA health care provider	11,467 (13)
Authorize sharing information with my VA health care team	8851 (10)
Authorize sharing information with other people (eg, family, caregiver)	7584 (9)
Other	6573 (7)
More Web-based educational programs	5396 (6)
Join a Web-based forum to discuss health issues with other veterans	3831 (4)
FY2008	**92,160**
View my upcoming appointments	79,892 (87)
Schedule or change my appointments	68,395 (74)
Look at information in my VA medical record	67,714 (73)
Web-based, secure communication with my doctor	58,878 (64)
Checking that different medications I take are safe when used together	45,986 (50)
Reminders of preventive care I need (eg, shots, cancer screening)	34,707 (38)
Notification of new content or features on the site	32,418 (35)
Advance check-in for my VA clinic visits	31,863 (35)
Monthly email newsletter	24,186 (26)
Share information that I have stored in My Health*e*Vet with other people	23,088 (25)
Advanced directive (eg, living will, durable power of attorney)	20,418 (22)
Educational programs	18,800 (20)
Information about the quality of VA health care	11,231 (12)
Other	8791 (10)

Table 4. My Health*e*Vet history and feature enhancement milestones. DoD: Department of Defense. EHR: electronic health record. VA: Department of Veterans Affairs.

Year	Milestone
1999	• My Health*e*Vet Pilot at 9 VA Medical Centers
2003	• National My Health*e*Vet Portal deployed
2004	• New user registration module deployed • Expansion of self-entered data modules
2005	• Prescription (Rx) Refill requests • Additional self-entered modules
2006	• In Person Authentication to Upgrade to Premium Account
2007	• Account Activity History • Forgot User ID and Password Support • Upgraded Health Calendar
2008	• Secure Messaging deployed for voluntary provider use • Master Veteran Index synchronization
2009	• VA Wellness Reminders
2010	• VA Blue Button Feature (Download My Data)
2011	• VA Appointments • VA Allergies • VA Chemistry and Hematology Lab Results • DoD Military Service Information • Display Rx Medication Name
2012	• Secure Messaging with all VA primary care providers • VA Immunizations • Veterans with DoD log-on credential can use to log in to portal (single sign on) • Social media content promotion
2013	• Expansion of VA EHR data in VA Blue Button Report (eg, VA Notes, VA Radiology Reports, Pathology Reports, Microbiology Lab Results, etc) • Basic VA Health Summary added • Hold Periods reduced from 7 to 3 calendar days • HealtheLiving Assessment (health risk appraisal) • Veterans Health Library
2014	• Ability to send Secure Messaging attachments • Migration to cloud environment for system stability, scalability, and performance • Log-in enhancements • Display medication images in pharmacy module
2015	• Secure Messaging Workload Credit • Rx Refill Shipment Email Notification • VA Appointment Email Reminders • Save Secure Messaging Progress Notes to VA EHR • Subscribe to My Health*e*Vet Newsletter
2016	• Content Management System deployed • Incremental Redesign: homepage dashboard navigation • Session time out warning
2017	• Enhanced VA Health Summary with Surgical and Clinical Procedure Notes • Incremental Redesign theme deployment • VA Medical Images and Reports Pilot • Personalized Veteran's Benefits Handbook • Appointment scheduling

Figure 2. Veteran comments about hold periods.



Why does it take so long for results of lab work, radiology, notes from Drs to show up? It can take a week or more. My Dr already called with the results yesterday but I still can't see it here. Also saw GI Dr 3 days ago and not notes here. I wish we could access our information sooner.

Eliminate wait period to view VA Notes, Results, etc. Once the provider has entered the note or viewed the results of test, they should be made available for viewing by the veteran.

Aligning Content With Patient-Suggested Topics of Interest

Periodically, an open-ended question is added to the survey asking users about topics of interest for portal content, such as feature articles to ensure that content is directly aligned with veterans' needs and preferences. An editorial calendar is created to provide articles throughout the year focused on these topics. Topics are also highlighted in a subscription-based monthly electronic newsletter that was developed in 2015 as a user-desired additional service (see Table 3), with more than 500,000 subscribers in FY2017. Examples of topic clusters for August 2009 (N=1809), August 2013 (N=3300), and August 2017 (N=1189) are shown in Figure 2. In 2017, the top user-suggested topics included Health ("general health, age concerns, pre-existing medical issues"), Diabetes ("articles on diabetes and feet or hand or finger neuropathy"), Care ("information on special health care programs for specific conditions"), and Agent Orange ("need more information on Agent Orange exposure and health issues").

Prospective Functional Design of a New Feature

Veteran feedback has also driven the functional design of new features. One feature that is currently being developed is the ability for the users to assign a delegate who can access their account. For example, a spouse or caregiver who may be assisting a veteran patient in managing his or her health. In October 2014, VA convened key stakeholders and subject matter experts to define the business requirements for this feature; however, there was a lack of consensus on a key functional requirement: whether "read-only" access should allow or restrict a delegate's ability to also print and download data. Using the CXA survey, veterans were asked "If you approve read access for another person to help you manage your personal health information, what would you want that person to be able to do?"

Of those veterans with a preference to delegate read access to another person, 75% (8194/11,006) would want such access to include print and download capability, while 14% (1541/11,006) would want a delegate to be able to read or view their information on the screen, but not print or download it. With this direct veteran input on desired functional design, requirements were prospectively aligned with user preferences. Data were also collected to assess patient preferences regarding delegating access to health information [43], use of My

Health*e*Vet to transfer information [44], how veterans with non-VA providers use the Blue Button feature to share information with their non-VA providers [45], and the veteran experiences with access to their VA Notes [36].

Website Redesign and Satisfaction Trends

Analysis of CXA data over the course of the last decade has been an integral part of the recent My Health*e*Vet website redesign initiative by enabling a deeper understanding of the end users and their tasks and goals, in keeping with UCD principles. As shown in Table 5, while 75% (75,241/100,617), 24% (23,923/100,617), and 18% (17,899/100,617) users in FY2008 accessed the portal to request a prescription refill, view their medication history, and look up information about a medication, respectively, user goals and tasks in FY2017 have shifted and expanded. Although prescription refill requests remained a predominant task (53,193/100,555, 53%), users also accessed the portal to view their VA Appointments (38,664/100,555, 38%), communicate with their health care team using Secure Messaging (28,952/100,555, 29%), track the delivery status of their medication refills (23,884/100,555, 24%), view their lab or test results (19,382/100,555, 19%), and access their VA health records (11,966/100,555, 12%). An important goal of the culminating website redesign was to improve navigation and usability for these specific core features, and the overall customer satisfaction index score was used as a performance indicator.

Historical customer satisfaction trends are shown in Figure 3. From October 2007 to October 2015, the aggregate average CXA score was 74, based on 945,480 completed surveys. The average for the 12 months that followed was stable at 76 (N=139,934). While multiple factors impacted customer satisfaction over the last decade, including a period of system performance issues in 2014 that was resolved by improving system architecture, the overall trend toward greater customer satisfaction is evident.

In October 2016, as part of an incremental website redesign, a dashboard was added to the portal home page to enhance user access to the core features (Figure 4).

As anticipated, the introduction of changes to the website resulted in an initial decrease in satisfaction (72), followed by satisfaction recovery (75), and subsequent increase to a new high of 79 (Figure 5). A similar pattern was observed with the deployment of additional website redesign changes in September 2017. Satisfaction initially decreased (77), but then recovered to previous levels (79). Satisfaction continued to increase in January 2018 (80).

Table 5. User-specified goals and tasks. FY: fiscal year. N/A: not applicable. VA: Department of Veterans Affairs.

Reason for visit or goal trying to accomplish[a]	FY2017 (N=100,555), n (%)	FY2008 (N=100,617), n (%)
Request a prescription refill	53,193 (53)	75,241 (75)
View my VA Appointments	38,664 (38)	N/A
Use Secure Messaging to communicate with my VA health care team	28,952 (29)	N/A
Track the status of my prescription refill delivery	27,516 (27)	N/A
View my medication history	23,884 (24)	23,923 (24)
View my lab or other test results	19,382 (19)	N/A
Access my VA health records or Blue Button or VA Health Summary	11,966 (12)	N/A
View my VA Notes (written by my health care team)	11,058 (11)	N/A
Look up information about a health condition or medication	9393 (9)	N/A
Learn more about features that are available	9149 (9)	N/A
Look up information about a medication	N/A	17,899 (18)
Find information about VA benefits	9111 (9)	6246 (6)
Enter or keep track of personal information	5695 (6)	14,507 (14)
Other	5101 (5)	9198 (9)
Enter or keep track of personal health care information (eg, blood pressure)	3202 (3)	13,125 (13)
Use the Veterans Health Library (Research a health condition)	2648 (3)	6367 (6)
Enter information about my non-VA medications or supplements	2288 (2)	N/A
Find a VA facility	1646 (2)	2206 (2)
Complete a HealtheLiving Assessment	1533 (2)	N/A

[a]Multiple categories may be selected.

Figure 3. Open-ended comment clusters for topics of interest.

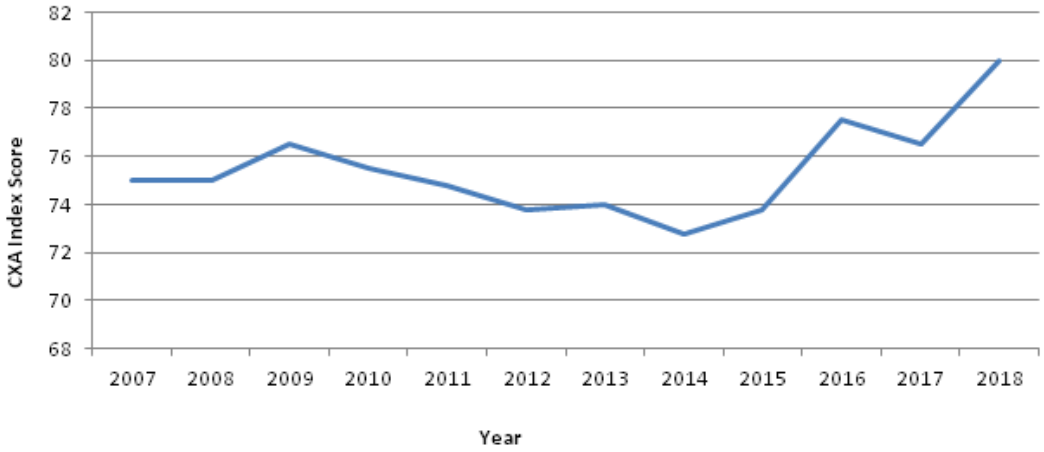
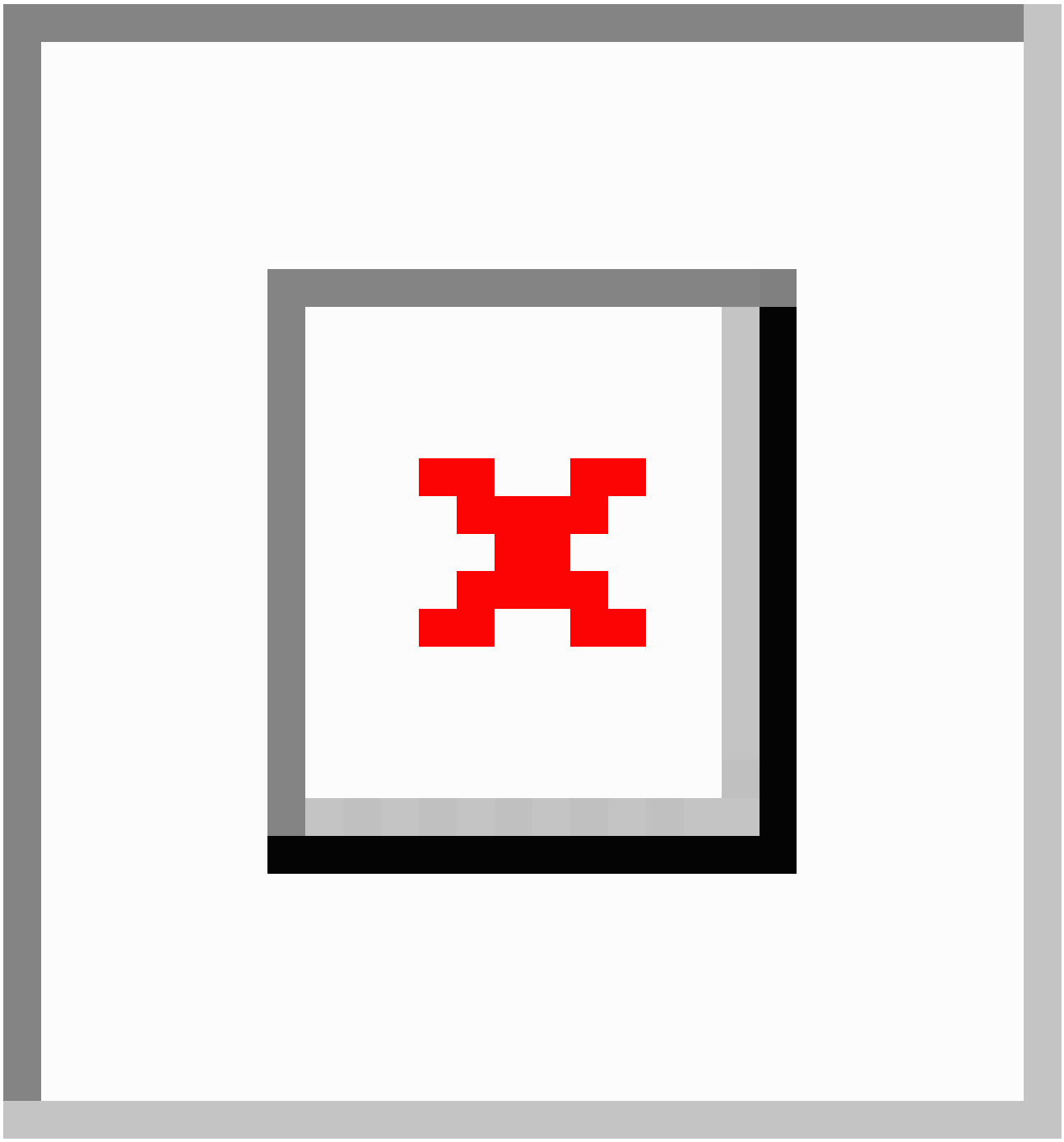


Figure 4. Historical customer satisfaction trends. CXA: customer experience analytics.

Figure 5. Incremental changes to My HealtheVet home page.

Figure 6. Customer experience analytics customer satisfaction index 15-month trend.



Discussion

Principal Findings

The literature on adoption and use of patient portals highlights the need for health care organizations to employ UCD approaches to ensure that portals align with end users' characteristics, needs, preferences, and goals, and, ultimately, help advance portal implementation. In its commitment to UCD, one method that VA has used to accomplish the above is a continuous survey to elicit a direct feedback from a random sample of veterans who use VA's patient portal, My Health*e*Vet. In combination with other methods, such as targeted research studies, the CXA survey has enabled a deeper understanding of portal users and directly informed changes in portal features, functions, policies, and processes. By incorporating the results of this systematic evaluation of the user experience into the portal redesign, VA aims to continue to enhance the ability of My Health*e*Vet to engage and activate veterans in managing their health.

Patient Portal Users

This study compared the characteristics and behaviors of users during the early period of patient portal implementation, 2008, with that of later adopters. This provided a trajectory of how portal use has evolved over a decade. Many aspects remained stable, while others showed clear trends toward portal adoption by populations believed less likely to use patient-facing health technologies. While only 13.4% (89,780/670,000) portal users in FY2008 were VA patients with a Premium account, by FY2017, this increased to 62.5% (2.5 million/4 million) users. Despite early assumptions about older users not adopting and using patient portals [46,47], leading to a gray digital divide [48], the VA experience reveals an increasingly elderly population of users. Within the veteran population, research has shown that VA patients tend to be older and more socioeconomically disadvantaged than veterans who do not rely on VA for care [49]. Although the survey indicates that the majority of users have one or more chronic health conditions and access the portal with increasing frequency, the survey results also suggest a trend toward those with less internet ability and better health also accessing the patient portal. This trend may be a result of the portal expanding the types of transactional services that users find convenient, based on direct veteran input. It also suggests that the portal is engaging a broader segment of the veteran population. Although the proportion of female veterans responding to the survey decreased slightly in FY2017 (from 9% to 7%), the overall population of female veterans was estimated to be 9.4% in 2015. However, only 22.4% used VA health care services [50], which is a key driver for accessing the patient portal. Portal users in FY2017 also tended to have completed higher levels of education than those in FY2008. This may be reflective of changes in the veteran population overall, with the enhanced provision of educational support programs for separating service members. Given that half of the survey respondents in FY2017 reported that they also use community non-VA care providers, VA will need to continue to develop tools that enable effective information sharing across settings of care. Portal functions that support consumer-mediated health information exchange are currently in early field testing [51]. These patient portal user trends align with similar trends for the VA patient population overall in terms of gender (91% male), age (median age of male VA patients, 64 years), and increasing use of VA education benefits [52].

Incremental Portal Redesign

Based on user self-report about goals and tasks, a significant redesign of the website was undertaken to enhance navigation to the features aligned with the most common user tasks and to decrease the number of steps to accomplish these. After an initial period of satisfaction decline, anticipated due to the phenomenon of change aversion [53], the satisfaction index recovered and increased. Once users adjusted to the change, they were more satisfied with the new design as measured using the CXA satisfaction index. Looking ahead, there are additional improvements and enhancements that will be important to address.

Limitations

It is important to note that the results of the CXA survey reflect the characteristics and perspectives of a random sample of portal users who are invited and opt to participate in the survey and may not be fully generalizable to the larger population. More broadly, the respondent sample represents patient portal users; therefore, other methods are also needed to elicit input from veterans who are not portal users to understand their characteristics and preferences and identify barriers that may exist to system access and use. VA is in the process of adding questions to its patient experience survey, administered to veterans who had a recent medical encounter, to help fill this gap, and ongoing research about veteran preferences for digital tools and services provides complementary insights [32,33]. There may also be data that were not collected in the survey that could be important. Since the survey is anonymous, there is no opportunity to follow up with respondents for more information or clarification. Despite limitations inherent to an anonymous survey, it has the benefit of enabling a continuous flow of direct feedback. While the findings from our case study may not be fully generalizable to other patient populations, the principle of using agile approaches to employ UCD has potential to be a promising implementation strategy for other health care organizations.

Conclusions

By leveraging UCD principles, VA has continued to enhance its patient portal and supported its continued implementation, achieving significant growth in adoption and use over the last decade. While quantitative and qualitative research studies are an important component of patient portal evaluation, more agile methods are also needed to complement formal research efforts. As illustrated through this case study, we have found the ongoing administration of a continuous voluntary Web-based survey as an efficient and effective way to capture veteran's voices about who they are, how they use the patient portal, what improvements are needed, and what additional services are desired. This approach, together with others intended to explore the perspectives of veterans who are not portal users, will help ensure that VA's health information technology services are

developed and enhanced to optimize the benefits to all VA patients. With impending changes to VA's EHR platform, capturing veteran's voices is more crucial than ever. More broadly, developing patient portals as an effective patient engagement strategy will require that UCD principles are employed to foster adoption and sustained use. In an era of finite resources, leveraging the "voice-of-the-customer" techniques helps ensure that the portal continues to meet patients' needs in ways that enhance full participation in their own health care.

Acknowledgments

The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. KMN was affiliated with the Department of Veterans Affairs during the time of this study but recently retired and is now an independent consultant specializing in Sociology, Communication, and Consumer Health Informatics.

Authors' Contributions

KMN was responsible for data collection. All authors (KMN, CLT, DMK, and TPH) contributed to the drafting of the manuscript and revising critically for intellectual content.

Conflicts of Interest

None declared.

References

1. Barello S, Graffigna G, Vegni E. Patient engagement as an emerging challenge for healthcare services: mapping the literature. Nurs Res Pract 2012;2012:905934 [FREE Full text] [doi: 10.1155/2012/905934] [Medline: 23213497]
2. Gephart S, Effken J. Online Journal of Nursing Informatics (OJNI). 2013. Using Health Information Technology to Engage Patients in their Care URL: http://ojni.org/issues/?p=2848 [accessed 2018-03-15] [WebCite Cache ID 6xv0k6wRS]
3. James J, Hibbard J, Agres T. Health Affairs. 2013. Health policy brief: patient engagement URL: https://www.healthaffairs.org/do/10.1377/hpb20130214.898775/full/ [accessed 2018-06-27] [WebCite Cache ID 70V6Ofvke]
4. Wilkins S. KevinMD.com. 2012 Jan 27. Patient engagement is the holy grail of health care URL: https://www.kevinmd.com/blog/2012/01/patient-engagement-holy-grail-health-care.html [accessed 2018-03-14] [WebCite Cache ID 6xv1fB2zh]
5. Chase D. Forbes. 2012 Sep. Patient engagement is the blockbuster drug of the century URL: https://www.forbes.com/sites/davechase/2012/09/09/patient-engagement-is-the-blockbuster-drug-of-the-century/ [accessed 2018-03-14] [WebCite Cache ID 6xv1wM7U1]
6. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito DA. Patient Portals as a Tool for Health Care Engagement: A Mixed-Method Study of Older Adults With Varying Levels of Health Literacy and Prior Patient Portal Use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]
7. Nazi K, Hogan T, Woods S. Consumer health informatics: engaging and empowering patients and families. In: Finnell J, Dixon B, editors. Clinical Informatics Study Guide. Switzerland: Springer International Publishing; 2016:459-500.
8. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
9. Sorondo B, Allen A, Fathima S, Bayleran J, Sabbagh I. Patient Portal as a Tool for Enhancing Patient Experience and Improving Quality of Care in Primary Care Practices. EGEMS (Wash DC) 2016;4(1):1262 [FREE Full text] [doi: 10.13063/2327-9214.1262] [Medline: 28203611]
10. Fraccaro P, Vigo M, Balatsoukas P, Buchan IE, Peek N, van DVSN. Patient Portal Adoption Rates: A Systematic Literature Review and Meta-Analysis. Stud Health Technol Inform 2017;245:79-83. [Medline: 29295056]
11. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
12. 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]
13. Tulu B, Trudel J, Strong DM, Johnson SA, Sundaresan D, Garber L. Patient Portals: An Underused Resource for Improving Patient Engagement. Chest 2016 Jan;149(1):272-277. [doi: 10.1378/chest.14-2559] [Medline: 26066707]
14. Centers for Medicare & Medicaid Services (CMS), HHS. Federal Register. 2015 Oct 16. Medicare and Medicaid programs: electronic health record incentive programs. Final rule URL: https://www.gpo.gov/fdsys/pkg/FR-2015-10-16/pdf/2015-25595.pdf [accessed 2018-06-28] [WebCite Cache ID 70ViCdppA]
15. Henry SL, Shen E, Ahuja A, Gould MK, Kanter MH. The Online Personal Action Plan: A Tool to Transform Patient-Enabled Preventive and Chronic Care. Am J Prev Med 2016 Jul;51(1):71-77. [doi: 10.1016/j.amepre.2015.11.014] [Medline: 26826751]
16. Lau M, Campbell H, Tang T, Thompson DJS, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]

17. 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]
18. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
19. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 01;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
20. OpenNotes. URL: https://www.opennotes.org/ [accessed 2018-05-12] [WebCite Cache ID 6zMPglGe7]
21. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 02;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
22. Dobscha SK, Denneson LM, Jacobson LE, Williams HB, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]
23. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What Patients Value About Reading Visit Notes: A Qualitative Inquiry of Patient Experiences With Their Health Information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
24. Leveille SG, Walker J, Ralston JD, Ross SE, Elmore JG, Delbanco T. Evaluating the impact of patients' online access to doctors' visit notes: designing and executing the OpenNotes project. BMC Med Inform Decis Mak 2012 Apr 13;12:32 [FREE Full text] [doi: 10.1186/1472-6947-12-32] [Medline: 22500560]
25. Wright E, Darer J, Tang X, Thompson J, Tusing L, Fossa A, et al. Sharing Physician Notes Through an Electronic Portal is Associated With Improved Medication Adherence: Quasi-Experimental Study. J Med Internet Res 2015 Oct 08;17(10):e226 [FREE Full text] [doi: 10.2196/jmir.4872] [Medline: 26449757]
26. Garrido T, Raymon B, Wheatley B. Health Affairs. 2016 Apr. Lessons learned from more than a decade in patient portals URL: https://www.healthaffairs.org/do/10.1377/hblog20160407.054362/full/ [accessed 2018-06-28] [WebCite Cache ID 70ViUCw5Q]
27. Veterans and Consumers Health Informatics Office. Department of Veterans Affairs. 2018. URL: https://connectedcare.va.gov/ [accessed 2018-06-28] [WebCite Cache ID 70VnehAAk]
28. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications; 2004.
29. Katz-Haaz R. Ten guidelines for user-centered web design. Usability Interface 1998;5(1):1-3.
30. Morales RM, Casper G, Brennan PF. Patient-centered design. The potential of user-centered design in personal health records. J AHIMA 2007 Apr;78(4):44-46. [Medline: 17455846]
31. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
32. Haun JN, Chavez M, Nazi K, Antinori N, Melillo C, Cotner BA, et al. Veterans' Preferences for Exchanging Information Using Veterans Affairs Health Information Technologies: Focus Group Results and Modeling Simulations. J Med Internet Res 2017 Oct 23;19(10):e359 [FREE Full text] [doi: 10.2196/jmir.8614] [Medline: 29061553]
33. Haun JN, Chavez M, Nazi KM, Antinori N. Developing a Health Information Technology Systems Matrix: A Qualitative Participatory Approach. J Med Internet Res 2016 Oct 06;18(10):e266 [FREE Full text] [doi: 10.2196/jmir.6499] [Medline: 27713112]
34. Hogan TP, Nazi KM, Luger TM, Amante DJ, Smith BM, Barker A, et al. Technology-assisted patient access to clinical information: an evaluation framework for blue button. JMIR Res Protoc 2014;3(1):e18 [FREE Full text] [doi: 10.2196/resprot.3290] [Medline: 24675395]
35. Nazi K. VAntage Point. 2010 Dec 08. VA's Blue Button: Empowering People with Their Data URL: https://www.blogs.va.gov/VAntage/866/vas-blue-button-empowering-people-with-their-data/ [accessed 2018-03-14] [WebCite Cache ID 6xv2X6xNS]
36. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
37. My HealtheVet. URL: https://www.myhealth.va.gov [accessed 2018-05-12] [WebCite Cache ID 6zLE6yNAn]
38. Nazi KM. Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. J Am Med Inform Assoc 2010;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]

39. Wood FB, Siegel ER, Feldman S, Love CB, Rodrigues D, Malamud M, et al. Web evaluation at the US National Institutes of Health: use of the American Customer Satisfaction Index online customer survey. J Med Internet Res 2008 Feb 15;10(1):e4 [FREE Full text] [doi: 10.2196/jmir.944] [Medline: 18276580]
40. Fornell C, Johnson MD, Anderson EW, Cha J, Bryant BE. Cornell University. 1996 Oct. The American Customer Satisfaction Index: nature, purpose, and findings URL: https://scholarship.sha.cornell.edu/cgi/viewcontent.cgi?article=1427&context=articles [accessed 2018-03-14] [WebCite Cache ID 6xv3L33fc]
41. Corbin J, Strauss A. Techniques and Procedures for Developing Grounded Theory. In: Basics of Qualitative Research. Thousand Oaks: Sage Publications; 2008.
42. Green J, Thorogood N. Qualitative Methods for Health Research. Thousand Oaks: Sage Publications; 2009.
43. Zulman DM, Nazi KM, Turvey CL, Wagner TH, Woods SS, An LC. Patient interest in sharing personal health record information: a web-based survey. Ann Intern Med 2011 Dec 20;155(12):805-810. [doi: 10.7326/0003-4819-155-12-201112200-00002] [Medline: 22184687]
44. Turvey C, Zulman D, Nazi K, Wakefield B, Woods S, Hogan T, et al. Transfer of information from personal health records: a survey of veterans using My HealtheVet. Telemed J E Health 2012 Mar;18(2):109-114. [doi: 10.1089/tmj.2011.0109] [Medline: 22304439]
45. Turvey C, Klein D, Fix G, Hogan TP, Woods S, Simon SR, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. J Am Med Inform Assoc 2014;21(4):657-663. [doi: 10.1136/amiajnl-2014-002723] [Medline: 24740865]
46. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
47. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. The Digital Divide and Patient Portals: Internet Access Explained Differences in Patient Portal Use for Secure Messaging by Age, Race, and Income. Med Care 2016 Dec;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
48. Morris A. E-literacy and the grey digital divide: a review with recommendations. JIL 2007 Dec 13;1(3):13-28. [doi: 10.11645/1.3.14]
49. RAND Corporation. Centers for Medicare and Medicaid (CMS). 2015 Sep 01. A product of the CMS Alliance to modernize healthcare federally funded research and development center URL: https://www.va.gov/opa/choiceact/documents/assessments/Assessment_A_Demographics.pdf [accessed 2018-03-14] [WebCite Cache ID 6xv4ifFsl]
50. National Center for Veterans Analysis and Statistics. Department of Veterans Affairs. 2017. Women Veterans Report: the past, present, and future of Women Veterans URL: https://www.va.gov/vetdata/docs/SpecialReports/Women_Veterans_2015_Final.pdf [accessed 2018-03-14] [WebCite Cache ID 6xv5EmKsF]
51. Williams C, Mostashari F, Mertz K, Hogin E, Atwal P. From the Office of the National Coordinator: the strategy for advancing the exchange of health information. Health Aff (Millwood) 2012 Mar;31(3):527-536 [FREE Full text] [doi: 10.1377/hlthaff.2011.1314] [Medline: 22392663]
52. National Center for Veterans Analysis and Statistics. Department of Veterans Affairs. 2017. VA Utilization Profile URL: https://www.va.gov/vetdata/docs/Quickfacts/VA_Utilization_Profile.pdf [accessed 2018-06-28] [WebCite Cache ID 70VmVA8g1]
53. Sedley A. GV Library. 2012 Apr 23. Change aversion: why users hate what you launched (and what to do about it) URL: https://library.gv.com/change-aversion-why-users-hate-what-you-launched-and-what-to-do-about-it-2fb94ce65766 [accessed 2018-03-14] [WebCite Cache ID 6xv5PQSHs]

Abbreviations

CXA: customer experience analytics
EHR: electronic health record
KP: Kaiser Permanente
UCD: user-centered design
VA: Department of Veterans Affairs

Edited by G Eysenbach; submitted 15.03.18; peer-reviewed by B Tulu, L Garvin, T Irizarry; comments to author 05.04.18; revised version received 14.05.18; accepted 16.06.18; published 10.07.18.

Please cite as:
Nazi KM, Turvey CL, Klein DM, Hogan TP
A Decade of Veteran Voices: Examining Patient Portal Enhancements Through the Lens of User-Centered Design
J Med Internet Res 2018;20(7):e10413
URL: http://www.jmir.org/2018/7/e10413/
doi:10.2196/10413
PMID:29991468

Original Paper

Relative Validity and Reproducibility of a New 44-Item Diet and Food Frequency Questionnaire Among Adults: Online Assessment

Aurélie Affret[1,2,3], PhD; Douae El Fatouhi[1,2,3], MSc; Courtney Dow[1,2,3], MPH; Emmanuelle Correia[1,2,3], BSc; Marie-Christine Boutron-Ruault[1,2,3], PhD; Guy Fagherazzi[1,2,3], PhD

[1]Inserm U1018, Center for Research in Epidemiology and Population Health, Villejuif, France

[2]Université Paris-Sud, Orsay, France

[3]Université de Versailles-Saint-Quentin-en-Yvelines, Versailles, France

Corresponding Author:
Marie-Christine Boutron-Ruault, PhD
Inserm U1018
Center for Research in Epidemiology and Population Health
114 Rue Edouard Vaillant
Villejuif, 94805
France
Phone: 33 33 142116466
Fax: 33 33 142114000
Email: marie-christine.boutron@gustaveroussy.fr

Abstract

Background: Dietary questionnaires currently available which can assess the habitual diet are timely, costly, or not adapted well to the modern diet; thus, there is a need for a shorter food frequency e-Questionnaire (FFeQ) adapted to Western diets, in order to properly estimate energy and macronutrient intakes or rank individuals according to food and nutrient intakes.

Objective: The aim of this study was to evaluate the relative validity and reproducibility of a 30-minute and 44-item FFeQ in a sample of adults obtained from the general population.

Methods: A sample of French adults was recruited through social media and an advertising campaign. A total of 223 volunteers completed the FFeQ twice at one-year intervals and were included in the reproducibility study. During that interval, 92 participants completed three-to-six 24-hour recalls and were included in the validity study. Nutrient and dietary intakes were computed for all validity and reproducibility participants. The level of agreement between the two methods was evaluated for nutrient and food group intakes using classification into quintiles of daily intake, correlation coefficients and Bland-Altman plots.

Results: For relative validity, correlation coefficients ranged from 0.09 to 0.88 (unadjusted correlation coefficients, median: 0.48) and 0.02 to 0.68 (deattenuated and energy adjusted correlation coefficients, median: 0.50) for food group and nutrient intakes, respectively. The median proportion of subjects classified into the same or adjacent quintile was 73% and 66% for food and nutrient intakes, respectively. Bland-Altman plots showed good agreement across the range of intakes. Regarding reproducibility, intraclass correlation coefficients ranged from 0.33 to 0.72 (median: 0.60) and 0.55 to 0.73 (median: 0.64), for food and nutrient intakes, respectively.

Conclusions: The FFeQ showed acceptable validity and reproducibility in a sample of adults based on their food and nutrient intakes. The FFeQ is a promising and low-cost tool that can be used in large-scale online epidemiological studies or clinical routines and could be integrated into evidence-based smartphone apps for assessing diet components.

(J Med Internet Res 2018;20(7):e227) doi:10.2196/jmir.9113

KEYWORDS

Short Food Frequency e-Questionnaire; Web-based; validity; reproducibility; online dietary assessment tool

Introduction

A healthy lifestyle, characterized by an adequate, balanced diet combined with regular physical activity, is one of the determinants for good health [1]. Moderate to strong associations between healthy dietary patterns and decreased risk of obesity and chronic diseases like cardiovascular disease, hypertension,

type 2 diabetes, and some cancers are strongly highlighted in the literature [2,3].

Among the many available tools that evaluate individual dietary intakes, the food frequency questionnaire (FFQ) has been frequently used in nutritional epidemiology studies since the 1980s [4-6]. Despite limitations (difficulties to estimate habitual intakes, memory bias, errors in perception of portion sizes, and use of restricted food lists), FFQs collect valuable information allowing researchers to assess the typical diet at a low cost and logistic burden, and it can be self-administered [5-7]. FFQs also allow researchers to accurately rank subjects according to their dietary and nutritional intake, which is important when comparing risks in various subgroups [6].

As previously described [8], a consortium of six cohort studies (E3N [9], E4N [10], CKD-REIN [11], i-Share [12], Elfe [13] and Psy-COH) was established to create a unique Food Frequency Questionnaire (FFQ) that could quickly assess the habitual diet in several populations: adults, the elderly, adolescents, students, patients with mental disorders, and patients with chronic kidney disease. It was decided that the questionnaire would be limited to 50 items to quickly assess the diet.

Existing FFQs of 50 items or less have not been proven suitable to quickly and accurately assess the diet in several French population subgroups: i) the FFQ by Vercambre et al [14] was developed for senior women and focused on specific nutrients for that population, ii) one third of the items of the FFQ developed by Giovannelli et al [15] were not valid and a food composition table was not available to study nutritional intakes, iii) the FFQ developed by Barrat et al [16] referred to intakes during the preceding week, and did not consider seasonal variability, and iv) two FFQs were developed solely in the context of cardiovascular prevention [17,18]. Therefore, the consortium decided to develop a new and unique FFQ (40 items) adapted to several population subgroups of interest. They agreed that some additional, specific questions (10 items maximum, as performed in the present study) could be added to the questionnaire to help estimate specific nutrients of interest or to obtain qualitative information about the dietary context, consistent with the targeted populations. In a pilot study, a paper version of the consortium FFQ was validated in a sample of patients with chronic kidney disease [8] and showed acceptable validity and reproducibility. Then, a web version of the questionnaire was developed by the consortium to meet the need of a food frequency e-Questionnaire (FFeQ) adapted to the diet of those in Western countries, able to accurately estimate energy and macronutrient intakes, and to rank individuals according to food and nutrient intakes.

Web-based dietary assessments provide a lot of advantages [19,20]: they have the potential to save time and financial resources, may be preferred by participants, and response quality can be improved directly by including cutoff values and alert messages in case of inconsistencies, abnormal or missing data. Several examples of 24-hour recall and Web-based FFQs already exist in the literature [21-26]. Because 24-hour recalls need to be repeated to assess the overall diet, Web-based FFQs might be more feasible in large-scale studies. However, most of the existing Web-based FFQs are long questionnaires and therefore time-consuming [21,24,25].

Before using a newly developed or modified FFQ, it must first be validated to be considered an acceptable method of dietary assessment [27]. The aim of the present study is to study the reproducibility of the online version of the newly developed FFQ (FFeQ) and evaluate its relative validity against 24-hour recalls in a sample of French adults.

Methods

Study Population and Design

According to Willett [6], the number of subjects necessary to conduct reproducibility and validity studies is approximately 110. Between January and February 2016, a national invitation to participate in the present reproducibility and validity study was advertised in the Inserm (*French* National *Institute* of Health and Medical Research) network and through the Inserm's Twitter and Facebook accounts. In total, 441 adults volunteered (from which 214 participants agreed to participate in the validity study), provided their informed consents, and were invited to complete the FFeQ. They were asked to complete the FFeQ twice, at a one-year interval, in February-April 2016 and again in February-April 2017. In total, 319 participants completed the FFeQ once (the 122 participants who failed to complete the first questionnaire were more likely to be women and live in the South and East of France and overseas). Two hundred twenty-nine participants completed it twice (participants who completed both questionnaires were more likely to be women and live in the South of France). Participants who under- or over- reported energy intake in one of the FFeQs, ie, were in the top and bottom 1% of the energy intake-to-energy requirement ratio distribution, were excluded as previously described [28]. Energy requirement was calculated as follows: Basal Metabolic Rate (BMR)* Physical Activity Level (the cutoff value of 1.55 for a minimal activity level was chosen [29]). BMR was computed based on sex, age, height, and weight, using the Schofield formula [30]. After exclusion, a total of 223 participants were included in the reproducibility study. Among them, 92 patients answered at least three (out of six) 24-hour recalls and were therefore included in the relative validity study (a flow diagram is presented in Multimedia Appendix 1). We decided to include participants with at least three recalls in the validity study to ensure i) reasonable intra-individual variations, ii) seasonal representation and iii) sufficient statistical power. Participants who completed at least three 24-hour recalls were more likely to be older than participants who completed the first questionnaire but did not complete three 24-hour recalls. They had higher energy intake and healthier dietary habits than their counterparts (data not tabulated).

From the 214 participants who volunteered for the validity study, only 130 were interviewed for the 24-hour recalls. The main reasons were contact difficulties as they were mostly active people, and that this study was conducted by only one dietitian which prevented us from interviewing several participants simultaneously.

Food Frequency Questionnaire

The food list for the FFeQ was developed by the investigators based on existing national food questionnaires [14-16,31-34] and data from the second national study of individual food intakes of French adults [35] to ensure that all food groups contributing to at least 5% of the mean energy, macronutrient, vitamin, or mineral intake of the French population were represented in the FFeQ, [16]. The FFeQ was self-administered online. The questionnaire asked participants to report their usual dietary intake over the past year. In epidemiological studies, one-year memory FFQ are mainly used because they assess long-term diets (diets tend to remain stable year on year) and season variability of intakes is considered [6].

The questionnaire was divided into two parts: The first part comprised 40 food groups. It quantified consumption by frequency (never or less than once a month, *x* times a day, *x* times a week or *x* times a month) and portion sizes per food group item. Photos previously validated [36] were directly integrated into the questionnaire to help participants estimate the consumed quantities of 21 food items (see Multimedia Appendices 2 and 3). Most of the time, there were three photos showing increasing portion sizes with five possible answers (less than the lowest portion, the lowest portion, an intermediate portion, the biggest portion, more than the biggest portion). For items not having a photo, participants were asked to quantify their consumption based on a standard portion size (typical household measurements like measuring spoons or standard units such as individual containers of yogurt).

The second part was specific to the study population. It was composed of 10 questions. Six were qualitative questions about eating habits (eg, meal frequency, socialization during meals, source of food supplies), and four questions were used to obtain nutritional data, of which two provided more detailed information about some food groups from the first part of the questionnaire (ie, fish and soft drinks).

In total, 44 items were used to obtain the nutritional data (see Multimedia Appendix 4). Daily intakes for each food group item were computed: frequencies were converted into numbers of servings per day and multiplied by the portion size. An ad hoc composition table was developed using data from the INCA2 French representative population survey (35) to estimate the percentage of contribution of each food included in a food group item. Nutritional data were then obtained using the French food composition database established by the French Data Centre on Food Quality (Ciqual, last updated in 2013) [37]. Besides nutritional and diet context information, information on sex, birth date, and anthropometric data was elicited. It also questioned participants about potential changes in their food habits during the past year due to specific situations (diet, pregnancy, a move, surgery or depression).

The FFeQ was adapted for laptops, tablets and smartphones. Compared to the paper version of the questionnaire [8], the online version presented here had a higher data quality thanks to alerts, restricted answers and automated checks. The design of the FFeQ consisted of one web page per food item and a pilot study previously demonstrated an average time of 30 minutes to complete the questionnaire.

24-Hour Recalls

The reference method used to compare results from the FFeQ consisted of six 24-hour recalls carried out every two months during the year between the first and the second FFeQ. Study participants were asked to recall all foods and beverages consumed on the previous day (due to logistics, data for Saturdays were collected on Mondays). Participants were not informed in advance of the day of the recall. To account for intra-individual variation (because dietary habits may differ according to weekdays or weekends and seasons), all days and all seasons were covered by the recalls as recommended [6] ie, days for recalls were randomly selected every two months per participant. Phone interviews were carried out by a trained dietitian who entered the data into the Nutrilog Software (v3.10b). These data were instantly converted into nutrient intakes by the software using the Ciqual food composition database [37]. A validated photo album of 42 foods [34] was previously e-mailed to the participants to help them quantify the amount of food consumed during the phone interview.

Statistical Analysis

We computed descriptive statistics (median and interquartile range) for nutrients and foods for both FFeQs and the average of the 24-hour recalls. Wilcoxon signed rank tests were performed to assess whether the mean ranks differed between groups.

Relative Validity

To study relative validity, data evaluated by the second FFeQ (FFeQ2) were compared with the mean of the 24-hour recalls, since both methods covered the same period. A list of concordance was established between food group items from the FFeQ and food items provided by 24-hour recalls. Few rarely consumed foods declared during the 24-hour recalls were not covered by the FFeQ items and were not taken into consideration.

Unadjusted Spearman correlation coefficients were calculated for food groups. Unadjusted and energy-adjusted Pearson correlation coefficients were calculated for nutrient intakes. Energy-adjusted coefficients, corrected for attenuation for within-person variation in the reference method (deattenuated coefficients) [6,38], were produced. Energy adjustment was performed using the residual method [6]. To improve the normal distribution, nutrient intakes were logarithmically transformed before analysis.

In terms of food group and nutrient intakes, we examined the level of agreement in ranking subjects between the two methods through cross-classification into quintiles. The percentage of participants classified in the lowest quintile in the FFeQ and the highest quintile in the 24-hour recalls (and vice versa) was studied. Because several food groups had a proportion of non-consumers >20%, we established three categories as follows: class=1 for null consumption; class=2 for consumption below or equal to the median value in consumers; class=3 for consumption over the median value in consumers. For food groups having a proportion of non-consumers <20%, subjects were classified into tertiles of consumption.

We evaluated agreement between the FFeQ and the 24-hour recalls performing Bland-Altman plots on energy-adjusted values [39-41]. Mean differences between the two assessment methods were plotted against the average estimation of the two methods. The 95% limit of agreement was calculated as the mean difference (SD 1.96).

Reproducibility

To evaluate reproducibility, data obtained from the first and second FFeQs (FFeQ1 and FFeQ2) were compared. For food groups, unadjusted Spearman correlation coefficients and intraclass correlation coefficients (ICC) were estimated. Unadjusted and energy-adjusted Pearson correlation coefficients as well as ICC were calculated for nutrient intakes. Nutrient intakes were logarithmically transformed before analysis, to improve the normal distribution. The level of agreement in ranking subjects between the two FFeQs (in terms of food group and nutrient intakes) was examined through cross-classification into quintiles. All statistical analyses were performed on SAS 9.4 (SAS Institute Inc, Cary, NC, USA). A *P* value <.05 was considered statistically significant.

Results

Baseline characteristics of the participants included in the relative validity and reproducibility studies are presented in Table 1. Participants included in both studies were mostly women (63.0%, 58 out of 92 participants and 74.9%, 67 out of 223 participants in the validity and reproducibility studies, respectively). Relative validity study participants were older than reproducibility study participants (47.7 years old, SD 14.9 vs 40.5 years old, SD 14.9) but the mean BMI was similar in both studies (23.5 kg/m^2, SD 4.2 in the reproducibility study). Participants lived in all regions of France, with higher proportions living in Paris and suburbs, and in the South. Most of the participants included in the relative validity study had complete data for six 24-hour recalls (73.9%, 68 out of 92 participants).

Relative Validity

Dietary intakes estimated by the FFeQ2 and the mean of the 24-hour recalls are presented in Table 2. Some food items such as “whole-grain pasta, rice, and wheat,” “legumes,” “milk” or “fruit” tended to be overestimated with the FFeQ whereas other food groups such as “raw vegetables,” “pizza, lasagna, and quiche,” “sausages and processed meat,” “cheese,” “sweet snacks, chocolate, and Danish pastries” or “alcoholic beverages excluding wine” were underestimated with the FFeQ2.

Unadjusted Spearman coefficients ranged from 0.09 (variety meats) to 0.88 (tea and herb teas), the median value being 0.48. Eight food groups had correlation coefficients below 0.3.

Table 1. Descriptive characteristics of the subjects included in the relative validity and reproducibility study.

Characteristic	Validity (n=92)	Reproducibility (n=223)
Sex (women), n (%)	58 (63.0)	167 (74.9)
Age (years), mean (SD)	47.7 (14.9)	40.5 (14.9)
Body mass index (kg/m^2), mean (SD)	23.6 (3.3)	23.5 (4.2)
Area of residence, n (%)		
South	15 (16.3)	60 (27.0)
West	10 (10.9)	19 (8.6)
North	1 (1.1)	8 (3.6)
East	7 (7.6)	13 (5.9)
Center	6 (6.5)	13 (5.9)
Overseas departments	0 (0)	3 (1.4)
Paris and suburbs	53 (57.6)	106 (47.7)
Number of 24-hour recall days, n (%)		
3	5 (5.4)	N/A[a]
4	4 (4.3)	N/A
5	15 (16.3)	N/A
6	68 (73.9)	N/A
Distribution of 24-hour recall days (average %)		
Weekday	67 (73.3)	N/A
Weekend	25 (26.7)	N/A
Autumn and winter	49 (53.0)	N/A
Spring and summer	43 (47.0)	N/A

[a]N/A: not applicable.

Table 2. Relative validity of the Food Frequency Questionnaire (FFeQ) for food groups (n=92). For food groups with a proportion of non-consumers >20%, tertiles and quintiles classifications were not performed. Instead, participants were classified as follows: class=1 for null consumption; class=2 for consumption below or equal to the median value in consumers, class=3 for consumption above the median value in consumers.

Food groups	Daily intakes[a]			FFeQ2 vs mean three to six 24-hour recalls			
	24-hour recalls, median (IQR)	FFeQ2, median (IQR)	FFeQ2-24-hour recalls, mean difference (SD)	Unadjusted Spearman correlation coefficients	Cross-classification of food group distribution, %		
					Subjects classified in same tertile	Subjects classified in same or adjacent quintile	Subjects classified in opposite quintiles
Whole-grain bread and substitutes[b]	18.7 (44.8)	10.7 (40.0)	-1.2 (54.1)	0.35	53	—[c]	—
White bread and substitutes	50.9 (57.7)	32.0 (69.3)	-2.2 (47.3)	0.63	54	78	1
Breakfast cereals[b]	0.0 (0.0)	0.0 (0.5)	1.9 (9.2)	0.68	79	—	—
Whole-grain pasta, rice and wheat[b,d]	0.0 (0.0)	2.3 (32.0)	24.2[d] (46.4)	0.14[e]	49	—	—
White pasta, rice and wheat	68.3 (82.5)	45.3 (90.7)	-11.3 (68.2)	0.52	49	71	1
Legumes[b]	0.0 (16.7)	16.7 (20.0)	8.8[d] (30.8)	0.26	27	—	—
French fries and other fried tubers[b]	0.0 (15.0)	8.7 (15.3)	2.3 (23.8)	0.35	40	—	—
Potatoes and other tubers (not fried)[b]	50.0 (64.6)	40.0 (53.3)	-4.8 (66.9)	0.26	45	—	—
Cooked vegetables	179.6 (163.0)	176.7 (166.7)	-8.8 (114.7)	0.49	50	68	1
Raw vegetables	63.1 (66.7)	34.4 (73.7)	-22.1[d] (51.7)	0.53	58	74	1
Pizza, lasagna and quiche[b]	27.9 (58.6)	14.2 (15.0)	-23.2[d] (40.3)	0.44	48	—	—
Sandwich, burgers and kebab[b]	0.0 (23.3)	0.0 (12.0)	-2.6 (18.2)	0.52	61	—	—
Fish fingers/breaded meat[b]	0.0 (0.0)	0.0 (3.3)	-4.1[d] (14.2)	0.27	64	—	—
Sausages and other processed meat[b]	30.0 (30.3)	11.3 (28.0)	-14.9[d] (29.5)	0.38	54	—	—
Poultry/rabbit[b]	24.2 (47.5)	20.0 (30.0)	4.1 (52.8)	0.45	41	—	—
Meat	43.3 (37.8)	26.3 (34.7)	-2.3 (44.4)	0.39	38	61	1
Variety meats[b]	0.0 (0.0)	0.0 (1.3)	1.7[d] (10.0)	0.09[e]	74	—	—
Eggs[b]	8.3 (20.0)	14.1 (21.2)	6.5[d] (26.8)	0.35	35	—	—
Fish[b]	15.8 (38.8)	13.3 (20.0)	-3.0 (26.2)	0.34	41	—	—
Seafood (excluding fish)[b]	0.0 (2.5)	0.0 (6.7)	-3.2 (19.1)	0.33	59	—	—
Milk[b]	0.0 (63.3)	0.0 (180.0)	70.6[d] (175.2)	0.73	68	—	—
Yogurt, white cheese, cottage cheese	96.3 (109.8)	125.0 (123.3)	28.0[d] (70.6)	0.75	60	88	0
Cream dessert[b]	0.0 (20.8)	2.0 (16.7)	5.1 (36.3)	0.40	53	—	—
Cheese	31.0 (30.6)	28.0 (18.0)	-7.1[d] (25.2)	0.51	52	68	1
Butter, fresh cream	4.2 (6.7)	10.0 (20.0)	10.1[d] (20.2)	0.66	51	79	0
Margarine, mayonnaise[b]	0.0 (0.4)	0.0 (0.8)	1.0[d] (4.4)	0.60	76	—	—
Olive oil	3.8 (6.3)	10.0 (14.7)	6.9[d] (10.0)	0.28	49	62	2
Rapeseed oil, walnut oil, mixed oil[b]	0.3 (2.0)	0.7 (2.7)	0.9[d] (4.5)	0.35	47	—	—
Sunflower oil, groundnut oil[b]	0.0 (0.0)	0.0 (0.7)	0.5 (2.8)	0.21	65	—	—
Salty snacks[b]	0.0 (8.3)	4.0 (8.0)	-0.1 (7.5)	0.41	43	—	—

Food groups	Daily intakes[a]			FFeQ2 vs mean three to six 24-hour recalls			
	24-hour recalls, median (IQR)	FFeQ2, median (IQR)	FFeQ2-24-hour recalls, mean difference (SD)	Unadjusted Spearman correlation coefficients	Cross-classification of food group distribution, %		
					Subjects classified in same tertile	Subjects classified in same or adjacent quintile	Subjects classified in opposite quintiles
Sweet snacks, chocolate, and Danish pastries[b]	46.7 (50.4)	13.3 (40.0)	-26.6[d] (61.0)	0.26	39	—	—
Fruit	202.8 (220.7)	351.0 (260.0)	122.3[d] (274.7)	0.67	61	77	1
Water	992.8 (746.7)	1550.0 (1200.0)	1134.5[d] (2410.6)	0.52	52	73	1
Coffee[b]	214.2 (303.3)	200.0 (380.0)	74.7 (303.4)	0.81	72	—	—
Tea and herb teas[b]	135.0 (557.1)	90.0 (400.0)	-3.1 (310.4)	0.88	71	—	—
Fruit juice[b]	50.0 (143.3)	50.0 (206.7)	29.8[d] (132.6)	0.61	58	—	—
Sweet beverages[b]	0.0 (2.1)	0.0 (0.0)	-11.5[d] (50.1)	0.73	79	—	—
Artificially-sweetened beverages[b]	0.0 (0.0)	0.0 (0.0)	10.8 (66.7)	0.55	88	—	—
Wine[b]	25.0 (90.0)	11.0 (40.5)	-16.4[d] (68.7)	0.81	67	—	—
Alcoholic beverages excluding wine[b]	13.3 (52.5)	2.5 (20.0)	-24.3[d] (65.5)	0.48	54	—	—

[a]Measured in grams (food) or milliliters (beverages).

[b]These food groups have a large proportion of non-consumers (>20%).

[c]Dashes indicate food groups that have a large proportion of nonconsumers (>20%). Classification into quintiles of consumption was not performed.

[d]The mean rank of the values of the three to six 24-hour recalls was significantly different to the mean rank of the values of the SFFeQ, according to Wilcoxon signed rank tests.

[e]Unadjusted Spearman correlation coefficients for which the statistical tests did not provide *P* values <.05.

The median proportion of participants classified in the same or adjacent quintiles of food group consumption by the FFeQ2, as well as by the mean of the 24-hour recalls, was 73%. The median proportion of participants classified in opposite quintiles was 1%. The median proportion of participants classified in the same tertile was 54%.

Mean macronutrient intakes estimated using the FFeQ2 did not differ in our study from those estimated in the 24-hour recalls (Table 3). Calcium and retinol intakes tended to be overestimated by the FFeQ (median 1113.7 mg/d, IQR 625.1 vs median 853.4 mg/d, IQR 322.5 and median 383.3 mg/d, IQR 393.4 vs median 0.0 µg/d, IQR 0.0, respectively) whereas alcohol and sodium intakes were underestimated by the FFeQ (median 2.0 g/d, IQR 6.9 vs median 4.7 mg/d, IQR 14.0 and median 2376.7 mg/d, IQR 945.7 vs median 2463.5 µg/d, IQR 1072.0, respectively).

Unadjusted correlation coefficients ranged from 0.08 (manganese and copper) to 0.77 (alcohol), with a 0.41 median value. Deattenuation mainly improved energy-adjusted correlation coefficients. Deattenuated energy-adjusted CC ranged from 0.05 (manganese) to 0.68 (potassium, carotene and vitamin C), with a 0.50 median value. A total of eight nutrients (sodium, magnesium, manganese, iron, copper, zinc, iodine and vitamin B12) had correlation coefficients lower than 0.3.

The median of percentages of participants classified in the same or adjacent quintiles of nutrient intakes by FFeQ2, and by the mean of the 24-hour recalls, was 66%. The median proportion of participants classified in opposite quintiles was 3%.

The Bland-Altman plot analysis graphs displayed good agreement between the two methods of estimation across the range of intake for energy (Figure 1), protein (Figure 2), carbohydrates (Figure 3), lipids (Figure 4), alcohol (Figure 5), cholesterol (Figure 6), sodium (Figure 7), and calcium intakes (Figure 8). For all 33 studied nutrients, the mean difference between methods (FFeQ2 vs means of 24-hour recalls) was close to zero for all levels of intake, except for calcium (data not shown). Across the range of intakes, calcium was systematically overestimated by the FFeQ2 which was consistent with the results displayed in Table 3. The percentage of points that were outside the limits of agreement ranged from 1.1% (zinc and iodine) to 7.6% (sugars, and vitamins D, B1, and B6), with a median value of 4.3%, which, theoretically is the percentage of values outside the mean (SD 1.96). Finally, the agreement did not differ between subjects with high and low intakes.

Table 3. Relative validity of the FFeQ for nutrients (n=92). Means and cross-classification were computed on crudes variables. All variables were log transformed before computing Pearson correlation coefficients to improve normality.

Nutrients	Daily intakes			FFeQ2 vs mean of three to six 24-hour recalls				
	24-hour recalls, median (IQR)	FFeQ2, median (IQR)	FFeQ2-24-hour recalls, mean (SD)	Pearson correlation coefficients[a]			Cross-classification of nutrient distribution, %	
				Unadjusted	Energy-adjusted[b]	Deattenuated[c]	Subjects classified in same or adjacent quintile	Subjects classified in opposite quintiles
Energy (kcal)	1882.2 (659.6)	1859.5 (819.8)	-77.4 (565.3)	0.47	N/A[d]	0.50[e]	66	3
Protein (g)	76.2 (22.6)	72.2 (34.1)	-3.5 (23.6)	0.57	0.47	0.52	72	2
Carbohydrates (g)	203.7 (74.6)	202.3 (98.0)	-1.3 (72.6)	0.44	0.49	0.54	71	3
Fat (g)	74.2 (28.4)	67.7 (34.8)	-3.6 (27.2)	0.47	0.55	0.61	59	3
SFA[f] (g)	26.9 (13.9)	25.8 (14.2)	-2.4[g] (10.5)	0.61	0.54	0.61	67	1
MUFA[h] (g)	24.2 (10.4)	26.3 (12.0)	3.1[g] (12.2)	0.35	0.46	0.53	58	7
PUFA[i] (g)	8.6 (3.7)	7.8 (3.6)	-0.1 (4.5)	0.32	0.48	0.55	58	4
Cholesterol (mg)	244.2 (133.6)	243.3 (130.2)	-11.4 (149.9)	0.46	0.31	0.39	65	5
Sugars (g)	79.8 (34.3)	82.5 (46.2)	3.6 (42.9)	0.28	0.39	0.45	71	8
Fiber (g)	21.6 (9.0)	20.9 (9.3)	-0.7 (7.9)	0.40	0.52	0.60	64	3
Alcohol (g)	4.7 (14.0)	2.0 (6.9)	-3.5[g] (8.5)	0.77[j]	—[k]	—[k]	88	1
Water (g)	2736.5 (868.5)	3182.4 (1744.0)	1310.8[g] (2654.0)	0.47	0.47	0.51	68	1
Sodium (mg)	2463.5 (1072.0)	2376.7 (945.7)	-402.5[g] (1483.9)	0.37	*0.05*	0.07	64	1
Magnesium (mg)	319.1 (104.6)	344.0 (194.9)	52.0[g] (142.4)	0.26	0.25	0.29	62	8
Phosphorus (mg)	1139.6 (334.3)	1114.7 (491.7)	1.6[g] (352.8)	0.50	0.40	0.45	67	2
Potassium (mg)	3099.1 (1044.0)	3091.6 (1555.0)	133.1 (955.3)	0.49	0.59	0.68	71	2
Calcium (mg)	853.4 (322.5)	1113.7 (625.1)	442.2[g] (553.1)	0.42	0.38	0.44	70	5
Manganese (mg)	2.8 (1.5)	11.5 (6.5)	9.3[g] (4.9)	*0.08*	*0.05*	0.05	59	8
Iron (mg)	9.4 (4.3)	10.6 (5.0)	1.1[g] (4.5)	0.28	0.23	0.27	62	3
Copper (mg)	1.4 (0.6)	2.2 (1.5)	0.8[g] (1.6)	*0.08*	*0.06*	0.07	57	7
Zinc (mg)	8.6 (2.9)	9.1 (4.7)	0.4 (5.0)	0.40	0.22	0.26	61	1
Iodine (µg)	120.2 (68.5)	122.1 (57.2)	-3.6 (69.1)	0.43	0.24	0.29	68	2
Retinol (µg)	0.0 (0.0)	383.3 (393.4)	506.6[g] (479.8)	0.11[j]	—[k]	—[k]	41	18
Carotene (µg)	3298.9 (2559.0)	3165.6 (2944.0)	-306.4 (2833.2)	0.52	0.54	0.68	77	3
Vitamin D (µg)	2.2 (1.5)	2.1 (1.2)	-0.4 (1.9)	0.32	0.23	0.31	61	3
Vitamin E (mg)	9.0 (4.3)	10.2 (5.1)	2.1[g] (5.4)	0.25	0.43	0.53	58	4
Vitamin C (mg)	118.6 (69.7)	124.7 (90.3)	13.3 (68.5)	0.54	0.56	0.68	78	2
Vitamin B1 (mg)	1.1 (0.4)	1.0 (0.4)	-0.1[g] (0.4)	0.31	0.33	0.41	66	4
Vitamin B2 (mg)	1.6 (0.6)	1.5 (0.8)	0.0 (0.6)	0.45	0.48	0.53	68	1
Vitamin B3 (mg)	16.5 (6.8)	15.6 (8.9)	0.0 (6.7)	0.49	0.49	0.58	73	4
Vitamin B5 (mg)	4.8 (2.0)	4.8 (2.4)	0.2 (1.8)	0.47	0.57	0.64	70	1
Vitamin B6 (mg)	1.7 (0.7)	1.6 (0.6)	-0.1 (0.7)	0.28	0.38	0.46	58	2
Vitamin B9 (µg)	305.1 (115.1)	321.0 (141.0)	26.9 (126.0)	0.36	0.49	0.58	64	1

Nutrients	Daily intakes			FFeQ2 vs mean of three to six 24-hour recalls				
	24-hour recalls, median (IQR)	FFeQ2, median (IQR)	FFeQ2-24-hour recalls, mean (SD)	Pearson correlation coefficients[a]			Cross-classification of nutrient distribution, %	
				Unadjusted	Energy-adjusted[b]	Deattenuated[c]	Subjects classified in same or adjacent quintile	Subjects classified in opposite quintiles
Vitamin B12 (µg)	3.3 (2.3)	5.6 (3.8)	2.1[g] (6.0)	0.31	0.21	0.26	65	7

[a]The statistical tests provided *P* values <.05 for each Pearson correlation coefficient, except for those in italics.

[b]Energy adjustment according to the residual method.

[c]Energy-adjusted and deattenuated Pearson correlation coefficient (corrected for within-person variation in the three to six 24-hour recalls).

[d]N/A: not applicable

[e]Unadjusted and de-attenuated Pearson correlation coefficient.

[f]SFA: saturated fatty acids.

[g]The mean rank of the values of the three to six 24-hr recalls was significantly different to the mean rank of the values of the SFFeQ, according to Wilcoxon signed rank tests. The statistical tests provided *P* values <.05 for each Pearson correlation coefficient, except for those in italics.

[h]MUFA: monounsaturated fatty acids.

[i]PUFA: polyunsaturated fatty acids.

[j]Spearman correlation coefficients were performed because normality was not respected.

[k]Normality was not respected. Energy-adjusted and deattenuated coefficients were not performed.

Figure 1. Bland-Altman plots related to energy. Difference in the daily intake of energy (crude variable) derived from the three to six 24-hour recalls and the short food frequency e-questionnaire (SFFeQ2) plotted against the corresponding mean daily intakes derived from the two methods. Solid lines represent mean difference, and dashed lines show lower and upper 95% limits of agreement (mean, SD 1.96; n=92).

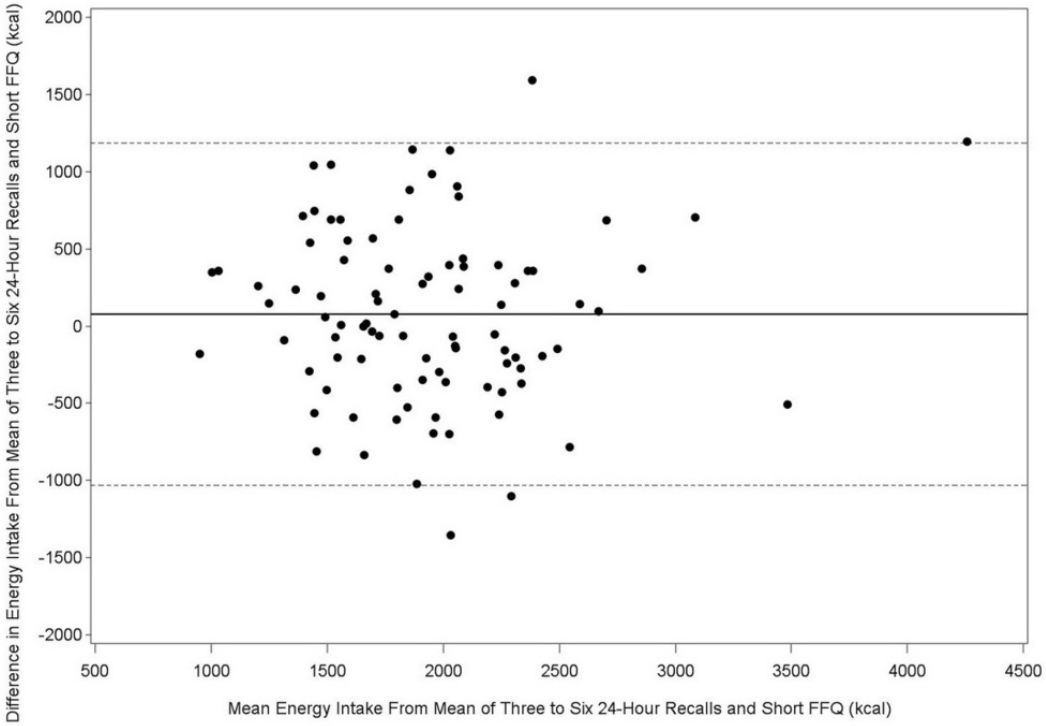


Figure 2. Bland-Altman plots related to protein. Difference in the daily intake of energy-adjusted protein derived from the three to six 24-hour recalls and the short food frequency e-questionnaire (SFFeQ2) plotted against the corresponding mean energy-adjusted daily intakes derived from the two methods. Solid lines represent mean difference, and dashed lines show lower and upper 95% limits of agreement (mean ± 1.96 SD) (n=92).

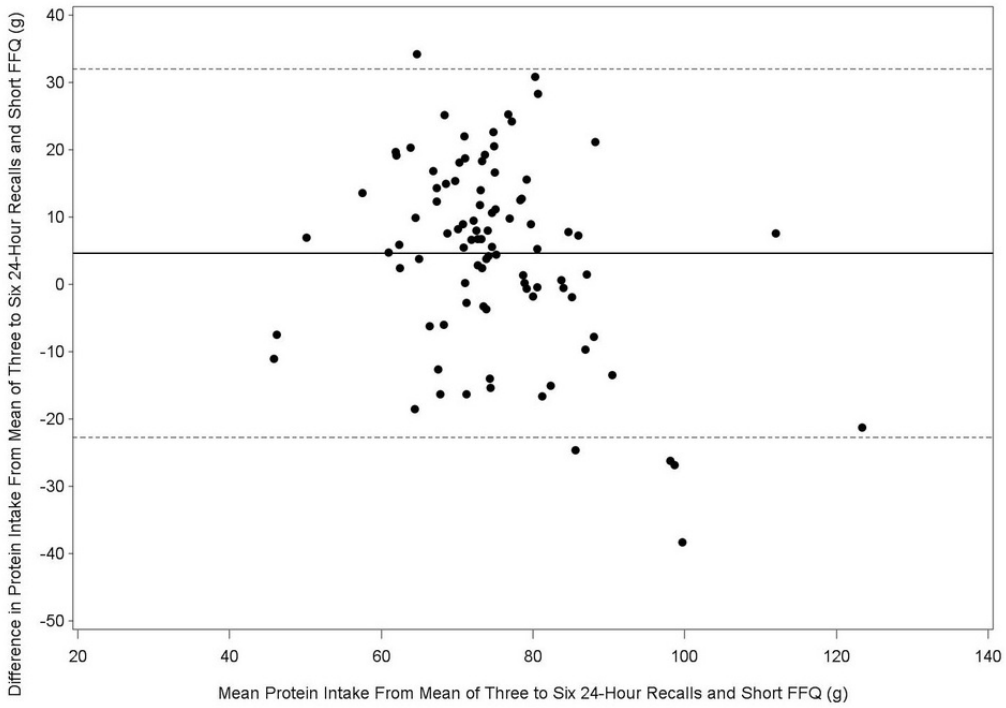


Figure 3. Bland-Altman plots related to carbohydrate. Difference in the daily intake of energy-adjusted carbohydrate derived from the three to six 24-hour recalls and the short food frequency e-questionnaire (SFFeQ2) plotted against the corresponding mean energy-adjusted daily intakes derived from the two methods. Solid lines represent mean difference, and dashed lines show lower and upper 95% limits of agreement (mean, SD 1.96; n=92).

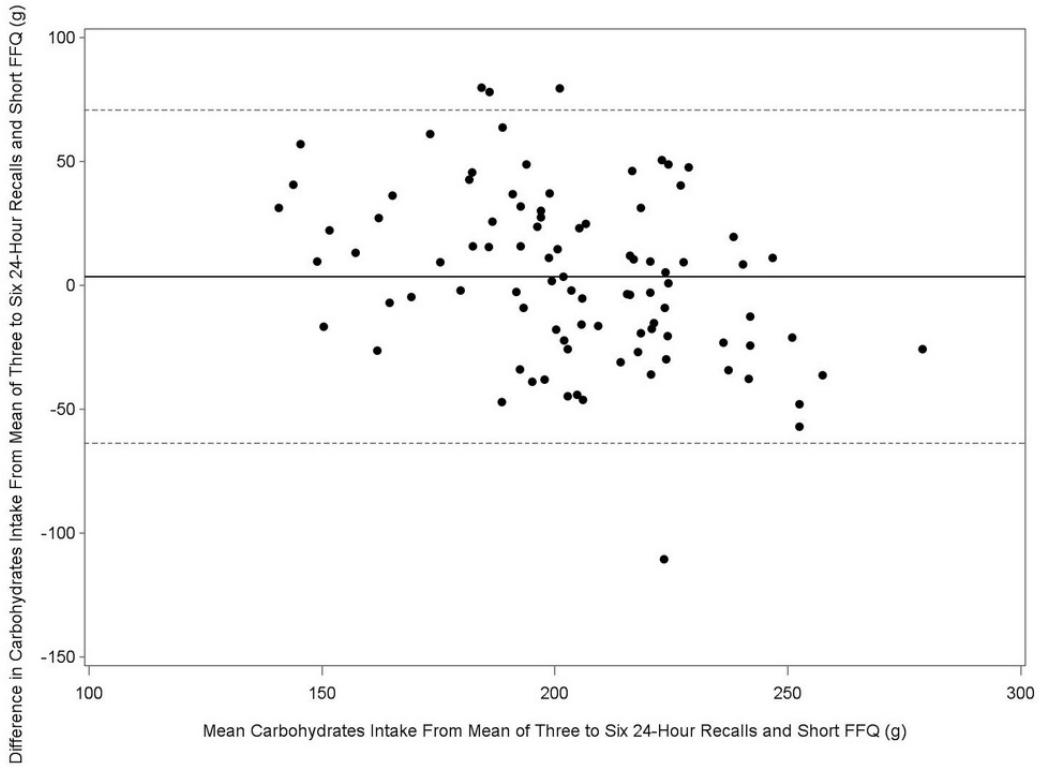


Figure 4. Bland-Altman plots related to lipid. Difference in the daily intake of energy-adjusted lipid derived from the three to six 24-hour recalls and the short food frequency e-questionnaire (SFFeQ2) plotted against the corresponding mean energy-adjusted daily intakes derived from the two methods. Solid lines represent mean difference, and dashed lines show lower and upper 95% limits of agreement (mean, SD 1.96; n=92).

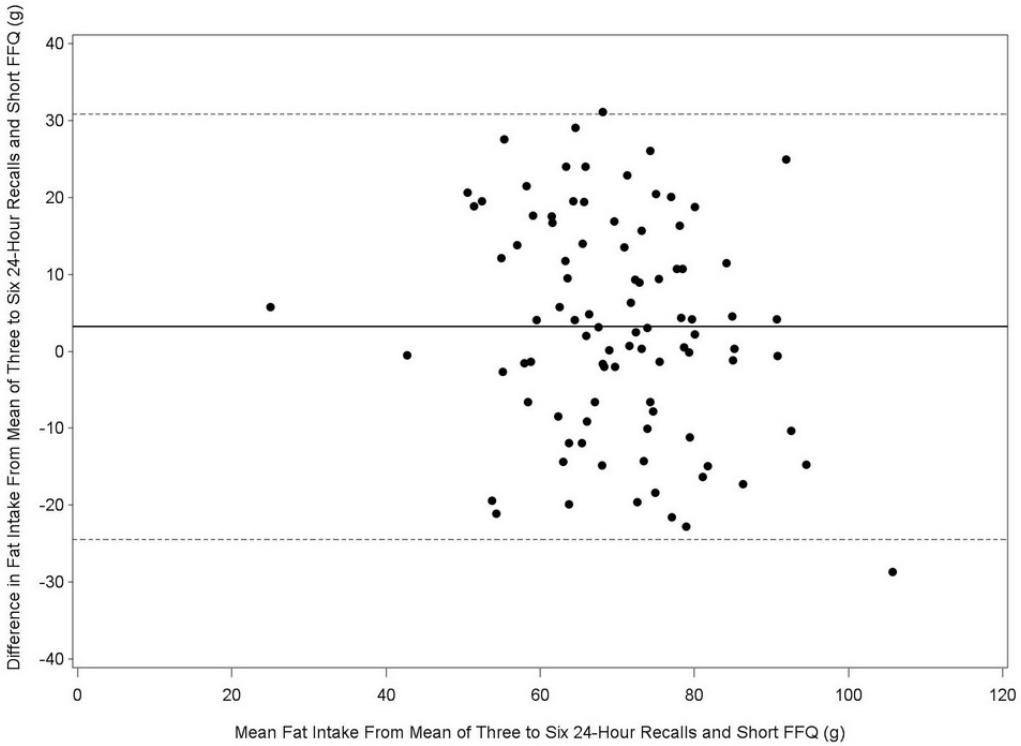


Figure 5. Bland-Altman plots related to alcohol. Difference in the daily intake of alcohol (crude variable) derived from the three to six 24-hour recalls and the short food frequency e-questionnaire (SFFeQ2) plotted against the corresponding mean daily intakes derived from the two methods. Solid lines represent mean difference, and dashed lines show lower and upper 95% limits of agreement (mean, SD 1.96; n=92).

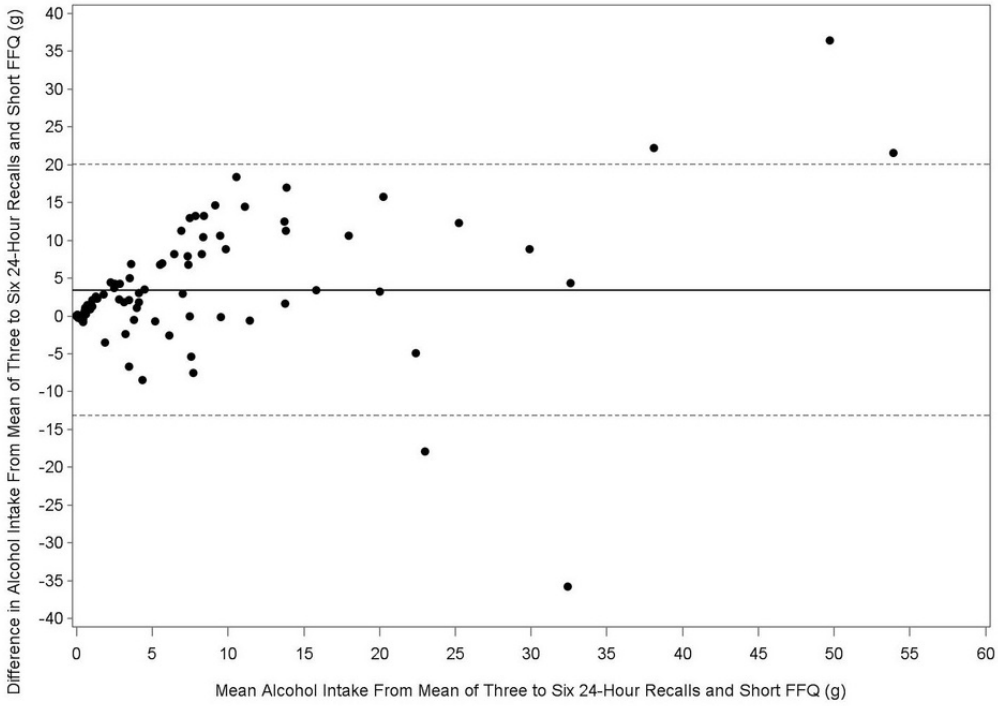


Figure 6. Bland-Altman plots related to cholesterol. Difference in the daily intake of energy-adjusted cholesterol derived from the three to six 24-hour recalls and the short food frequency e-questionnaire (SFFeQ2) plotted against the corresponding mean energy-adjusted daily intakes derived from the two methods. Solid lines represent mean difference, and dashed lines show lower and upper 95% limits of agreement (mean, SD 1.96; n=92).

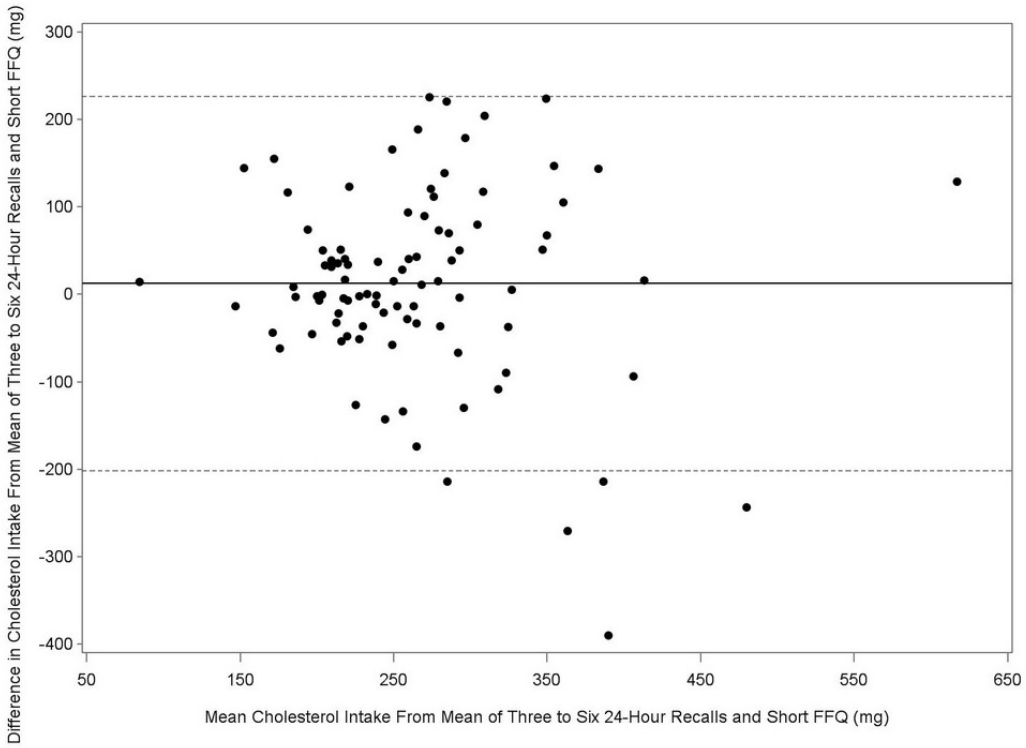


Figure 7. Bland-Altman plots related to sodium. Difference in the daily intake of energy-adjusted sodium derived from the three to six 24-hour recalls and the short food frequency e-questionnaire (SFFeQ2) plotted against the corresponding mean energy-adjusted daily intakes derived from the two methods. Solid lines represent mean difference, and dashed lines show lower and upper 95% limits of agreement (mean, SD 1.96; n=92).

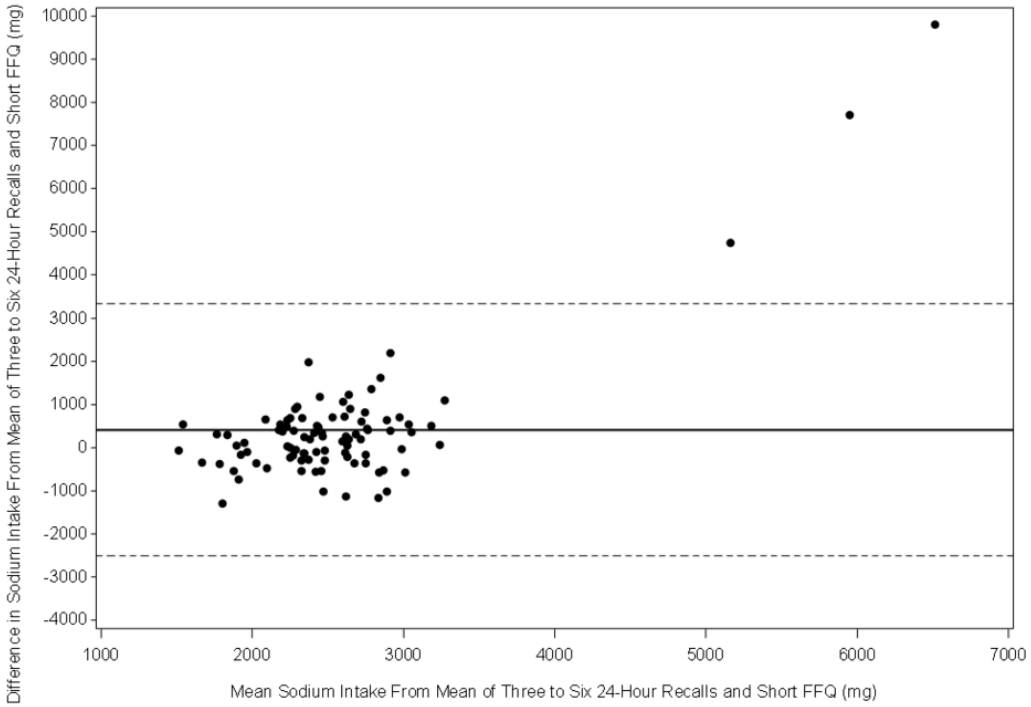
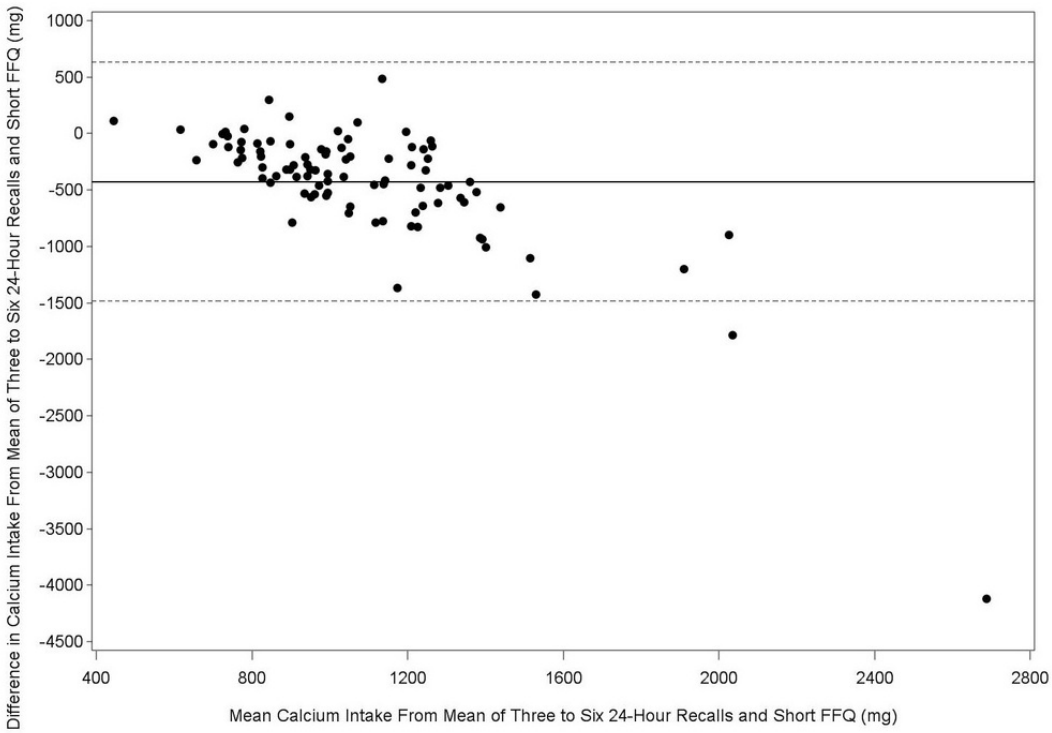


Figure 8. Bland-Altman plots related to calcium. Difference in the daily intake of energy-adjusted calcium derived from the three to six 24-hour recalls and the short food frequency e-questionnaire (SFFeQ2) plotted against the corresponding mean energy-adjusted daily intakes derived from the two methods. Solid lines represent mean difference, and dashed lines show lower and upper 95% limits of agreement (mean, SD 1.96; n=92).



Reproducibility

Absolute daily intakes of food groups were mostly comparable between the two FFeQs (see Multimedia Appendix 5). A rather large statistically significant decrease was observed for the "sweet beverages" and "artificially-sweetened beverages" between FFeQ1 and FFeQ2.

Unadjusted Spearman correlation coefficients ranged from 0.34 (sunflower and groundnut oils) to 0.90 (wine), with the median value being 0.65. Intraclass correlation coefficients ranged from 0.33 (sweet snacks, chocolate, and Danish pastries) to 0.72 (poultry or rabbit, fish, and fruit), with the median value being 0.60.

The median of percentages of subjects classified in the same or adjacent quintiles of food group consumption by both FFeQs was 80%. The median proportion of participants classified in opposite quintiles was 1%. The median percentage of subjects classified in the same tertile was 64%.

Absolute daily intake of energy and nutrients were comparable between the two FFeQs, although all nutrient intakes (excluding alcohol) showed a slight but statistically significant decrease between FFeQ1 and FFeQ2 (see Multimedia Appendix 6).

Crude correlation coefficients ranged from 0.58 (iron) to 0.89 (alcohol), with a 0.65 median value. Energy-adjusted Pearson correlation coefficients ranged from 0.54 (vitamin B1) to 0.77 (vitamin E), with a 0.65 median value. Intraclass correlation coefficients ranged from 0.55 (carbohydrates) to 0.73 (magnesium and manganese), the median value being 0.65.

The median proportion of subjects classified in the same or adjacent quintiles of nutrient intakes by both FFeQs was 79%. The median proportion of participants classified in opposite quintiles was 1%.

Discussion

Principal Findings

The present study investigated the relative validity and reproducibility of a new FFeQ, in a sample of French adults. The FFeQ was designed to accurately estimate energy intake and rank participants according to their dietary and nutrient intakes. The overall results indicate acceptable relative validity (for nutrient intakes, median correlation coefficient=0.50 and median proportion of subjects classified in the same or adjacent quintiles by the FFeQ2 and the 24-hour recalls=66%), and good reproducibility (for nutrient intakes, median correlation coefficient=0.65, and median proportion of subjects classified in the same or adjacent quintiles=79%). Our tool demonstrated an acceptable ability to rank participants for most nutrients and food groups, making it sufficiently informative when studying associations with health outcomes and when adjusting for nutritional intake in epidemiological and clinical studies [6,42]. It can also be used to derive dietary patterns using collected food data.

Because combinations of different assessment methods are becoming increasingly popular and can address several methodological limitations [43], it would be interesting, in further analyses to study under which circumstances a combination of the FFeQ with dietary recalls would be more efficient than either the FFeQ or the dietary recalls alone to address precision, power, and sample size, as it has been previously done [44].

Our FFeQ was composed of 50 items, of which 44 were used to obtain nutritional data. In their review, Cade et al reported that the number of items in FFQs published between 1980 and 1999 ranged from three to 350 items [45], with the median being 79 items. However, according to Willett [6], there is a rapidly decreasing marginal gain in information obtained with increasingly detailed questionnaires. Considering the relative validity and reproducibility of our FFeQ, it appears that the chosen 44 items were sufficient to assess the overall diet and to describe major food and nutrient intakes. The remaining six items of the online FFeQ will help provide meaningful insights as they could be used for qualitative studies or to stratify statistical analyses according to eating habits like meal frequency, socialization during meals or source of food supplies.

Relative Validity

Recovery biomarkers are considered gold-standard measures to validate self-reported intakes. However, because of their costliness and because only a few recovery biomarkers are currently known, they are rarely used in validity studies [43]. When selecting a reference method to validate a tool, errors of both methods must be as independent as possible [46]. Even if correlated errors related to memory, perception of serving sizes and social desirability exist between FFQs and 24-hour recalls, multiple 24-hour recalls have often been considered as the best feasible reference method [47]. Here, to study the FFeQ relative validity, three to six 24-hour recalls were used as the reference method.

Our study showed acceptable relative validity for food (correlation coefficients median and range: 0.51 [0.09-0.88]) and nutrients (correlation coefficients median and range: 0.40 [0.05-0.68]), and our results were comparable to those from other studies [14,21,32,33,46,48-52]. In the NIH–AARP study, validity coefficients for energy-adjusted nutrients ranged from 0.36 to 0.76 [46].

For validation studies, it has been suggested that correlation coefficients should be ≥ 0.3, preferably over 0.4, and optimally in the range of 0.5-0.7 [6,45,53]. In our study, only eight food groups out of 40 and eight nutrients out of 34 had correlation coefficients lower than 0.3.

For food groups, the lowest correlation coefficients were found for foods that are not consumed regularly such as “variety meats,” “whole-grain pasta, rice and wheat,” “legumes,” and “sunflower oil and groundnut oil”. Such findings on rarely consumed foods have been previously reported [32-34]. Low correlation coefficients were also observed for mixed items such as the “sweet snacks, chocolate, and Danish pastries” item. Due to the number of foods included in these groups, consumption may be difficult to estimate. Finally, a low correlation coefficient was observed for the “potatoes and other tubers” item. The estimation of its consumption may have been difficult due to the large consumption of mixed dishes including potatoes in France. Underestimation of water during the 24-hour recall has previously been reported [34] arguing that even if its consumption is systematically asked, it is easily forgotten.

Because of social desirability [42,54], food groups such as “alcoholic beverages” and “fruit” may have been under- and overestimated respectively. Over-reporting of fruit and vegetable intake by subjects seeking social approval is a common bias [55]. As previously reported [56,57], correlation coefficients were lower for vegetables than for fruit. According to Wakai [57], it may be partly because the frequency of fruit consumption is easier to report than vegetables because fruit is more often consumed raw whereas vegetables are more frequently part of cooked dishes and therefore not integrally recalled. Furthermore, fruit is frequently consumed as a single food item and comes in natural or typical units, whereas vegetables are often sliced or cut which makes them more difficult to quantify [56].

In our study, unadjusted correlation coefficients for macronutrients ranged from 0.47 to 0.57. Similar results were previously obtained for online FFQs (range: 0.06-0.68 [21,48,50], FFQs of 50 items max; range: 0.22-0.53 [14,51,52] and FFQs developed for adults; range: 0.29-0.61 [32,33,49]).

All macronutrients had correlation coefficients in the range of 0.5-0.7. Regarding the nutrients, the highest correlation coefficient was observed for alcohol (unadjusted correlation coefficient of 0.77). One of the lowest coefficients was observed for sodium (deattenuated energy adjusted correlation coefficient of 0.07). As reported in the pilot study (validation study of the paper version of the FFQ among patients with chronic kidney disease [8]), even though a question about salt added after food preparation was asked in the specific part of the questionnaire, it was still difficult to estimate its intake. However, when looking at individuals’ rankings, 64% of participants were classified in the same or adjacent quintile in terms of sodium intake when comparing the FFeQ2 and the 24-hour recalls.

After adjustment for energy some correlation coefficients were increased, and others were decreased. According to Willett et al [58], energy adjustment can increase the correlation coefficients when the variability of the nutrient intake is related to energy intake, or it can decrease when the variability of the nutrient is subject to systematic errors of under or overestimation of reported food consumption.

Despite some differences in estimations in both foods and nutrients by the questionnaire, agreement in classification was comparable to what other studies have shown [59] or slightly lower than shown in other studies [14,32]. Our results were close to the recommended 70% [60]. The highest level of participants classified in opposite quintiles was observed for retinol (18%). One of the main sources of retinol are variety meats, which were rarely consumed and for which consumption was probably difficult to evaluate with only three to six 24-hour recalls.

Reproducibility

Our study showed acceptable reproducibility for most foods (ICC range: 0.33-0.72, median 0.60) and nutrients (ICC range: 0.55-0.73, median 0.65). Our findings were comparable to prior reported correlation coefficients for reproducibility [25,49,61]. An important factor influencing reproducibility is the period between the two questionnaires. We adopted a one-year time interval which is long but frequently used and reported as acceptable [45,62,63]. However, we cannot exclude that some dietary changes may have occurred during the period. The

reproducibility observed here may therefore be lower than the true value.

As previously reported in the literature, a slight decrease in food and nutrient intakes was observed between FFeQ1 and FFeQ2 [16,25,32,64,65]. Due to the completion of the 24-hour recalls and the FFeQ1, a learning effect may explain this trend [66]. In favor of this hypothesis, several authors found that the second FFQ, which indicated reduced nutrient intake, was more valid than the first one when compared to 24-hour recalls [32,65,67].

According to a review, correlation coefficients of 0.5 to 0.7 between two administrations are commonly reported [45]. In our study, 75% and 100% of the studied food groups and nutrients had correlation coefficients ≥0.5. Not one food group had a correlation coefficient lower than 0.3.

Agreement in classification was very good (median of 80% and 79% for food and nutrient intakes respectively). For all nutrient intakes 74% to 94% of participants were classified in the same or adjacent quintile.

Strengths and Limitations

The current work has some limitations. People involved in the current study were volunteers. Volunteers may be more health-conscious, pay more attention to their diets than the average population, and therefore provide more accurate responses to questionnaires. However, subjects participating in observational epidemiological or clinical studies that are likely to use this tool in the future are also volunteers.

Our work has several strengths. The tool we developed was easy to complete and not time consuming. The implementation of photographs helped the participants estimate the amounts of food consumed and it has previously been shown that the use of photographs improves the ability to report the true quantity of dietary intakes [68].

In addition, a total of 92 participants were included in our study. It is higher than the reported number in recent studies [21,69,70].

Here, we present relative validity results for the online FFQ in a sample of French adults. A paper version of the FFQ has previously been validated in a sample of patients with chronic kidney disease [8] and further validity studies will now be conducted in specific population subgroups (for example, adolescents or cancer survivors). One of the main strengths of the consortium is that we will have a unique tool (due to the shared 40 items in the first part of the FFQ), that is useful for the comparison between several populations. The questionnaire is now available for other epidemiological and clinical studies interested in assessing the habitual diet quickly. We validated a Web-based version of the FFQ which provides valuable insights: it enables an interactive interface for participants and improves the quality of answers by directly including cutoff values and messages of alert in case of inconsistent, abnormal, or missing data.

Conclusions

For most food groups and nutrients, the FFeQ showed acceptable relative validity and reproducibility in a sample of French adults. It appears to be valid to rank individuals based on their food and nutrient intakes and can now be used in large-scale epidemiological studies as well as in clinical routine to easily and quickly assess the habitual diet. Developing an evidence-based smartphone application from the FFeQ is the next step. This type of tool may further be used to monitor patients' nutrient intakes and provide them with instantaneous feedback and nutritional recommendations about their diets.

Acknowledgments

This work was supported by a grant from the French Research Institute in Public Health (IReSP) and the French multi-organism thematic Institutes in Public Health (ITMO Santé Publique). AA is supported by a doctoral grant from the French National Cancer Institute. GF is supported by the French Research Agency (ANR, Agence Nationale de la Recherche) via an "Investissement d'Avenir" grant (investment for the future grant, ANR-10-COHO-0006) that supports the E4N study. The IReSP, ITMO Santé Publique, and ANR had no role in the design, analysis or writing of this article. The authors are grateful to all participants for providing the data used in this relative validity and reproducibility study. We thank Maryvonne Niravong and Françoise Clavel-Chapelon for their advice and contributions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Flow diagram.

[JPG File, 110KB - jmir_v20i7e227_app1.JPG]

Multimedia Appendix 2

Extract from the SFFeQ, French and original version.

[JPG File, 76KB - jmir_v20i7e227_app2.jpg]

Multimedia Appendix 3

Extract from the SFFeQ, version translated in English.

[JPG File, 86KB - jmir_v20i7e227_app3.jpg]

Multimedia Appendix 4

SFFQ items used to obtain nutritional data (n=44).

[JPG File, 140KB - jmir_v20i7e227_app4.JPG]

Multimedia Appendix 5

Reproducibility of food group consumptions of the SFFeQ (n=223).

[JPG File, 239KB - jmir_v20i7e227_app5.JPG]

Multimedia Appendix 6

Reproducibility of nutrient intakes of the SFFeQ (n=223).

[JPG File, 167KB - jmir_v20i7e227_app6.JPG]

References

1. WHO | Nutrition Internet. WHO URL: http://www.who.int/topics/nutrition/en/ [accessed 2018-05-30] [WebCite Cache ID 6znEVKjVc]
2. WCRF / AICR. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective 2009 URL: http://www.aicr.org/assets/docs/pdf/reports/Second_Expert_Report.pdf [accessed 2018-05-30] [WebCite Cache ID 6znEYjBCO]
3. Dietary GAC. Scientific Report of the 2015 Dietary Guidelines Advisory Committee. Scientific Advisory Report of the 2015 Dietary Guidelines Advisory Committee to the Secretary of HealthHuman Servicesthe Secretary of Agriculture: U.S. Department of Agriculture, Agricultural Research Service, Washington D. C; 2015. URL: https://health.gov/dietaryguidelines/2015-scientific-report/pdfs/scientific-report-of-the-2015-dietary-guidelines-advisory-committee.pdf [accessed 2018-05-30] [WebCite Cache ID 6znErxPU3]
4. Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. Adv Nutr 2015 Jan;6(1):5-18 [FREE Full text] [doi: 10.3945/an.114.007492] [Medline: 25593140]
5. Subar AF, Kushi LH, Lerman JL, Freedman LS. Invited Commentary: The Contribution to the Field of Nutritional Epidemiology of the Landmark 1985 Publication by Willett et al. Am J Epidemiol 2017 Jun 01;185(11):1124-1129. [doi: 10.1093/aje/kwx072] [Medline: 28535308]
6. Willett W. Nutritional Epidemiology. Third Edition. Oxford, New York: Oxford University Press; 2012:978.
7. Shim J, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]
8. Affret A, Wagner S, El FD, Dow C, Correia E, Niravong M, CKD-REIN study investigators, et al. Validity and reproducibility of a short food frequency questionnaire among patients with chronic kidney disease. BMC Nephrol 2017 Sep 15;18(1):297 [FREE Full text] [doi: 10.1186/s12882-017-0695-2] [Medline: 28915857]
9. Clavel-Chapelon F, van LMJ, Giubout C, Niravong MY, Goulard H, Le CC, et al. E3N, a French cohort study on cancer risk factors. E3N Group. Etude Epidémiologique auprès de femmes de l'Education Nationale. Eur J Cancer Prev 1997 Oct;6(5):473-478. [Medline: 9466118]
10. Clavel-Chapelon F, E3N Study Group. Cohort Profile: The French E3N Cohort Study. Int J Epidemiol 2015 Jun;44(3):801-809. [doi: 10.1093/ije/dyu184] [Medline: 25212479]
11. Stengel B, Combe C, Jacquelinet C, Briançon S, Fouque D, Laville M, et al. [The French Chronic Kidney Disease-Renal Epidemiology and Information Network (CKD-REIN) cohort study: To better understand chronic kidney disease]. Nephrol Ther 2016 Apr;12 Suppl 1:S49-S56. [doi: 10.1016/j.nephro.2016.01.005] [Medline: 26976057]
12. Montagni I, Guichard E, Carpenet C, Tzourio C, Kurth T. Screen time exposure and reporting of headaches in young adults: A cross-sectional study. Cephalalgia 2016 Oct;36(11):1020-1027. [doi: 10.1177/0333102415620286] [Medline: 26634831]
13. Vandentorren S, Bois C, Pirus C, Sarter H, Salines G, Leridon H, Elfe team. Rationales, design and recruitment for the Elfe longitudinal study. BMC Pediatr 2009 Sep 21;9:58 [FREE Full text] [doi: 10.1186/1471-2431-9-58] [Medline: 19772571]
14. Vercambre M, Boutron-Ruault M, Niravong M, Berr C, Clavel-Chapelon F, Ragusa S. Performance of a short dietary questionnaire to assess nutrient intake using regression-based weights. Public Health Nutr 2009 Apr;12(4):547-552. [doi: 10.1017/S1368980008002310] [Medline: 18457600]

15. Giovannelli J, Dallongeville J, Wagner A, Bongard V, Laillet B, Marecaux N, et al. Validation of a short, qualitative food frequency questionnaire in French adults participating in the MONA LISA-NUT study 2005-2007. J Acad Nutr Diet 2014 Apr;114(4):552-561. [doi: 10.1016/j.jand.2013.07.002] [Medline: 24083967]
16. Barrat E, Aubineau N, Maillot M, Derbord E, Barthes P, Lescuyer J, et al. Repeatability and relative validity of a quantitative food-frequency questionnaire among French adults. Food Nutr Res Internet 2012 Oct 30;56:18472.
17. Béliard S, Coudert M, Valéro R, Charbonnier L, Duchêne E, Allaert FA, et al. Validation of a short food frequency questionnaire to evaluate nutritional lifestyles in hypercholesterolemic patients. Ann Endocrinol (Paris) 2012 Dec;73(6):523-529. [doi: 10.1016/j.ando.2012.09.006] [Medline: 23141865]
18. Laviolle B, Froger-Bompas C, Guillo P, Sevestre A, Letellier C, Pouchard M, et al. Relative validity and reproducibility of a 14-item semi-quantitative food frequency questionnaire for cardiovascular prevention. Eur J Cardiovasc Prev Rehabil 2005 Dec;12(6):587-595. [Medline: 16319550]
19. Benedik E, Koroušić SB, Simčič M, Rogelj I, Bratanič B, Ding EL, et al. Comparison of paper- and web-based dietary records: a pilot study. Ann Nutr Metab 2014;64(2):156-166. [doi: 10.1159/000363336] [Medline: 25116257]
20. Illner A, Freisling H, Boeing H, Huybrechts I, Crispim SP, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. Int J Epidemiol 2012 Aug;41(4):1187-1203 [FREE Full text] [doi: 10.1093/ije/dys105] [Medline: 22933652]
21. Bentzen SMR, Knudsen VK, Christiensen T, Ewers B. Relative validity of a web-based food frequency questionnaire for patients with type 1 and type 2 diabetes in Denmark. Nutr Diabetes 2016 Sep 26;6(9):e232 [FREE Full text] [doi: 10.1038/nutd.2016.40] [Medline: 27669176]
22. Arab L, Tseng C, Ang A, Jardack P. Validity of a multipass, web-based, 24-hour self-administered recall for assessment of total energy intake in blacks and whites. Am J Epidemiol 2011 Dec 01;174(11):1256-1265 [FREE Full text] [doi: 10.1093/aje/kwr224] [Medline: 22021561]
23. Carter MC, Albar SA, Morris MA, Mulla UZ, Hancock N, Evans CE, et al. Development of a UK Online 24-h Dietary Assessment Tool: myfood24. Nutrients 2015 Jun;7(6):4016-4032 [FREE Full text] [doi: 10.3390/nu7064016] [Medline: 26024292]
24. Forster H, Fallaize R, Gallagher C, O'Donovan CB, Woolhead C, Walsh MC, et al. Online dietary intake estimation: the Food4Me food frequency questionnaire. J Med Internet Res 2014;16(6):e150 [FREE Full text] [doi: 10.2196/jmir.3105] [Medline: 24911957]
25. Labonté M, Cyr A, Baril-Gravel L, Royer M, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. Eur J Clin Nutr 2012 Feb;66(2):166-173. [doi: 10.1038/ejcn.2011.163] [Medline: 21934698]
26. Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, Bingley C, et al. The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. J Acad Nutr Diet 2012 Aug;112(8):1134-1137 [FREE Full text] [doi: 10.1016/j.jand.2012.04.016] [Medline: 22704899]
27. Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. Public Health Nutr 2002 Dec;5(6A):915-923. [doi: 10.1079/PHN2002383] [Medline: 12633516]
28. Cottet V, Touvier M, Fournier A, Touillaud MS, Lafay L, Clavel-Chapelon F, et al. Postmenopausal breast cancer risk and dietary patterns in the E3N-EPIC prospective cohort study. Am J Epidemiol 2009 Nov 15;170(10):1257-1267. [doi: 10.1093/aje/kwp257] [Medline: 19828509]
29. FAO/WHO/ONU. Energy and Protein Requirements. Report of a Join Expert Consultation URL: http://apps.who.int/iris/bitstream/10665/39527/1/WHO_TRS_724_%28chp1-chp6%29.pdf [accessed 2018-05-30] [WebCite Cache ID 6znFTaHYa]
30. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. Hum Nutr Clin Nutr 1985;39 Suppl 1:5-41. [Medline: 4044297]
31. Bonifacj C, Gerber M, Scali J, Daures JP. Comparison of dietary assessment methods in a southern French population: use of weighed records, estimated-diet records and a food-frequency questionnaire. Eur J Clin Nutr 1997 Apr;51(4):217-231. [Medline: 9104572]
32. Deschamps V, de LB, Lafay L, Borys J, Charles MA, Romon M. Reproducibility and relative validity of a food-frequency questionnaire among French adults and adolescents. Eur J Clin Nutr 2009 Feb;63(2):282-291 [FREE Full text] [doi: 10.1038/sj.ejcn.1602914] [Medline: 17882132]
33. Kesse-Guyot E, Castetbon K, Touvier M, Hercberg S, Galan P. Relative validity and reproducibility of a food frequency questionnaire designed for French adults. Ann Nutr Metab 2010;57(3-4):153-162. [doi: 10.1159/000321680] [Medline: 21079389]
34. van LMJ, Lucas F, Clavel F, Slimani N, Villeminot S. Relative validity and reproducibility of a French dietary history questionnaire. Int J Epidemiol 1997;26 Suppl 1:S128-S136 [FREE Full text] [Medline: 9126541]
35. AFSSA (Agence Française de Sécurité Sanitaire des Aliments). 2009. INCA2 - Etude Individuelle Nationale des Consommations Alimentaires 2 URL: https://www.anses.fr/fr/system/files/PASER-Ra-INCA2.pdf [accessed 2018-05-30] [WebCite Cache ID 6znFVh3P5]

36. Lucas F, Niravong M, Villeminot S, Kaaks R, Clavel-Chapelon F. Estimation of food portion size using photographs: validity, strengths, weaknesses and recommendations. J Hum Nutr Diet 1995 Feb;8(1):65-74. [doi: 10.1111/j.1365-277X.1995.tb00296.x]
37. Ciqual. URL: https://ciqual.anses.fr/ [accessed 2018-05-30] [WebCite Cache ID 6znGmMBzj]
38. Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. Am J Epidemiol 1988 Feb;127(2):377-386. [Medline: 3337089]
39. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
40. Bland JM, Altman DG. Agreed statistics: measurement method comparison. Anesthesiology 2012 Jan;116(1):182-185. [doi: 10.1097/ALN.0b013e31823d7784] [Medline: 22129533]
41. Altman D. Practical Statistics for Medical Research. London: CRC Press; 1990:978.
42. Hébert JR, Hurley TG, Steck SE, Miller DR, Tabung FK, Peterson KE, et al. Considering the value of dietary assessment data in informing nutrition-related health policy. Adv Nutr 2014 Jul;5(4):447-455 [FREE Full text] [doi: 10.3945/an.114.006189] [Medline: 25022993]
43. Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. F1000Res 2017;6:926 [FREE Full text] [doi: 10.12688/f1000research.10703.1] [Medline: 28690835]
44. Carroll RJ, Midthune D, Subar AF, Shumakovich M, Freedman LS, Thompson FE, et al. Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. Am J Epidemiol 2012 Feb 15;175(4):340-347 [FREE Full text] [doi: 10.1093/aje/kwr317] [Medline: 22273536]
45. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires - a review. Public Health Nutr 2002 Aug;5(4):567-587. [doi: 10.1079/PHN2001318] [Medline: 12186666]
46. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985 Jul;122(1):51-65. [Medline: 4014201]
47. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. J Natl Cancer Inst 2011 Jul 20;103(14):1086-1092 [FREE Full text] [doi: 10.1093/jnci/djr189] [Medline: 21653922]
48. Apovian CM, Murphy MC, Cullum-Dugan D, Lin P, Gilbert KM, Coffman G, et al. Validation of a web-based dietary questionnaire designed for the DASH (dietary approaches to stop hypertension) diet: the DASH online questionnaire. Public Health Nutr 2010 May;13(5):615-622 [FREE Full text] [doi: 10.1017/S1368980009991996] [Medline: 19912673]
49. Bassett JK, English DR, Fahey MT, Forbes AB, Gurrin LC, Simpson JA, et al. Validity and calibration of the FFQ used in the Melbourne Collaborative Cohort Study. Public Health Nutr 2016 Sep;19(13):2357-2368. [doi: 10.1017/S1368980016000690] [Medline: 27075344]
50. Christensen SE, Möller E, Bonn SE, Ploner A, Wright A, Sjölander A, et al. Two New Meal- and Web-Based Interactive Food Frequency Questionnaires: Validation of Energy and Macronutrient Intake. J Med Internet Res Internet 2013 Jun 2013 Jun 05;15:e109-e124.
51. Mannato LW, Pereira TSS, Velasquez-Melendez G, Cardoso LDO, Benseñor IM, Molina MDCB. Comparison of a short version of the Food Frequency Questionnaire with its long version--a cross-sectional analysis in the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). Sao Paulo Med J 2015;133(5):414-420 [FREE Full text] [doi: 10.1590/1516-3180.2014.00533004] [Medline: 26648430]
52. Verger EO, Armstrong P, Nielsen T, Chakaroun R, Aron-Wisnewsky J, Gøbel RJ, MetaCardis Consortium. Dietary Assessment in the MetaCardis Study: Development and Relative Validity of an Online Food Frequency Questionnaire. J Acad Nutr Diet 2017 Jun;117(6):878-888. [doi: 10.1016/j.jand.2016.10.030] [Medline: 28024800]
53. Masson LF, McNeill G, Tomany JO, Simpson JA, Peace HS, Wei L, et al. Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. Public Health Nutr 2003 May;6(3):313-321. [doi: 10.1079/PHN2002429] [Medline: 12740081]
54. Hebert JR, Ma Y, Clemow L, Ockene IS, Saperia G, Stanek EJ, et al. Gender differences in social desirability and social approval bias in dietary self-report. Am J Epidemiol 1997 Dec 15;146(12):1046-1055. [Medline: 9420529]
55. Hebert JR, Clemow L, Pbert L, Ockene IS, Ockene JK. Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. Int J Epidemiol 1995 Apr;24(2):389-398. [Medline: 7635601]
56. De KW, Dekkers A, Van VV, Ottevaere C, Van OH, De HS, et al. Relative validity of a short qualitative food frequency questionnaire for use in food consumption surveys. Eur J Public Health 2013 Oct;23(5):737-742. [doi: 10.1093/eurpub/cks096] [Medline: 22843611]
57. Wakai K. A review of food frequency questionnaires developed and validated in Japan. J Epidemiol 2009;19(1):1-11 [FREE Full text] [Medline: 19164867]
58. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr 1997 Apr;65(4 Suppl):1220S-1228S; discussion 1229S. [Medline: 9094926]
59. Sluik D, Bezemer R, Sierksma A, Feskens E. Alcoholic Beverage Preference and Dietary Habits: A Systematic Literature Review. Crit Rev Food Sci Nutr 2016 Oct 25;56(14):2370-2382. [doi: 10.1080/10408398.2013.841118] [Medline: 25674684]

60. Pietinen P, Hartman AM, Haapa E, Räsänen L, Haapakoski J, Palmgren J, et al. Reproducibility and validity of dietary assessment instruments. II. A qualitative food frequency questionnaire. Am J Epidemiol 1988 Sep;128(3):667-676. [Medline: 2843041]
61. Cantin J, Latour E, Ferland-Verry R, Morales SS, Lambert J, Faraj M, et al. Validity and reproducibility of a food frequency questionnaire focused on the Mediterranean diet for the Quebec population. Nutr Metab Cardiovasc Dis 2016 Feb;26(2):154-161. [doi: 10.1016/j.numecd.2015.11.003] [Medline: 26719221]
62. Kaaks R, Slimani N, Riboli E. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S26-S36. [Medline: 9126531]
63. Molag ML, de VJHM, Ocké MC, Dagnelie PC, van DBPA, Jansen MCJF, et al. Design characteristics of food frequency questionnaires in relation to their validity. Am J Epidemiol 2007 Dec 15;166(12):1468-1478. [doi: 10.1093/aje/kwm236] [Medline: 17881382]
64. Fernández-Ballart JD, Piñol JL, Zazpe I, Corella D, Carrasco P, Toledo E, et al. Relative validity of a semi-quantitative food-frequency questionnaire in an elderly Mediterranean population of Spain. Br J Nutr 2010 Jun;103(12):1808-1816. [doi: 10.1017/S0007114509993837] [Medline: 20102675]
65. Marques-Vidal P, Ross A, Wynn E, Rezzi S, Paccaud F, Decarli B. Reproducibility and relative validity of a food-frequency questionnaire for French-speaking Swiss adults. Food Nutr Res Internet 2011 May 2011 May 06;55:5905.
66. Wong SS, Boushey CJ, Novotny R, Gustafson DR. Evaluation of a computerized food frequency questionnaire to estimate calcium intake of Asian, Hispanic, and non-Hispanic white youth. J Am Diet Assoc 2008 Mar;108(3):539-543. [doi: 10.1016/j.jada.2007.12.006] [Medline: 18313437]
67. Malekshah AF, Kimiagar M, Saadatian-Elahi M, Pourshams A, Nouraie M, Goglani G, et al. Validity and reliability of a new food frequency questionnaire compared to 24 h recalls and biochemical measurements: pilot phase of Golestan cohort study of esophageal cancer. Eur J Clin Nutr 2006 Aug;60(8):971-977. [doi: 10.1038/sj.ejcn.1602407] [Medline: 16465196]
68. Liu X, Wang X, Lin S, Song Q, Lao X, Yu IT. Reproducibility and Validity of a Food Frequency Questionnaire for Assessing Dietary Consumption via the Dietary Pattern Method in a Chinese Rural Population. PLoS One 2015;10(7):e0134627 [FREE Full text] [doi: 10.1371/journal.pone.0134627] [Medline: 26230275]
69. Buscemi S, Rosafio G, Vasto S, Massenti FM, Grosso G, Galvano F, et al. Validation of a food frequency questionnaire for use in Italian adults living in Sicily. Int J Food Sci Nutr 2015;66(4):426-438. [doi: 10.3109/09637486.2015.1025718] [Medline: 25830946]
70. Fallaize R, Forster H, Macready AL, Walsh MC, Mathers JC, Brennan L, et al. Online dietary intake estimation: reproducibility and validity of the Food4Me food frequency questionnaire against a 4-day weighed food record. J Med Internet Res 2014;16(8):e190 [FREE Full text] [doi: 10.2196/jmir.3355] [Medline: 25113936]

Abbreviations

BMI: body mass index
BMR: basal metabolic rate
FFQ: Food Frequency Questionnaire
FFeQ: Food Frequency e-Questionnaire

Edited by G Eysenbach; submitted 05.10.17; peer-reviewed by M Boaz, P Wark, N Bragazzi, H Miller; comments to author 10.12.17; revised version received 01.02.18; accepted 18.02.18; published 05.07.18.

Please cite as:
Affret A, El Fatouhi D, Dow C, Correia E, Boutron-Ruault MC, Fagherazzi G
Relative Validity and Reproducibility of a New 44-Item Diet and Food Frequency Questionnaire Among Adults: Online Assessment
J Med Internet Res 2018;20(7):e227
URL: http://www.jmir.org/2018/7/e227/
doi:10.2196/jmir.9113
PMID:29980502

Original Paper

Investigating the Direct Impact of a Gamified Versus Nongamified Well-Being Intervention: An Exploratory Experiment

Saskia Marion Kelders[1,2], PhD; Marion Sommers-Spijkerman[1], MSc; Jochem Goldberg[1], MSc

[1]Center for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands
[2]Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa

Corresponding Author:
Saskia Marion Kelders, PhD
Center for eHealth and Wellbeing Research
Department of Psychology, Health and Technology
University of Twente
Drienerlolaan 5
Enschede,
Netherlands
Phone: 31 534896055
Email: s.m.kelders@utwente.nl

Abstract

Background: Gamification is a promising strategy to increase the effectiveness of Web-based mental health interventions by enhancing engagement. However, because most studies focus on the longer term effects of gamification (eg, effectiveness or adherence at the end of the intervention period), there is limited insight into how gamification may enhance engagement. Research implies that gamification has a direct impact at the time of use of the intervention, which changes the experience of the users, and thereby motivates users. However, it is unclear what this direct impact of gamification might be and how it can be measured.

Objective: The objective of this study was to explore the direct impact of gamification on behavioral, cognitive, and affective engagement in the context of a Web-based mental health intervention and to explore whether and how the different components of engagement are related.

Methods: A pilot (n=19) and a real-life (n=75) randomized between-groups experiment was carried out, where participants used a gamified or nongamified version of the same Web-based well-being intervention for a single session. Participants (68%, 64/94 female, mean age 23 years) were asked to use the intervention in one session for research purposes. Gamification elements included a map as visualization of the different lessons, a virtual guide, and badges. Later, behavioral, cognitive, and affective engagement were measured.

Results: The pilot experiment showed no differences between the gamified and nongamified intervention. However, in the real-life experiment, participants in the gamified intervention scored higher on cognitive engagement, that is, involvement (P=.02) and some elements of affective engagement, that is, flow as a combination of cognitive and affective engagement (P=.049), and the emotions ”interest” (P=.03) and “inspiration” (P=.009). Furthermore, the effect of gamification on cognitive engagement was mediated by the influence of gamification on specific positive emotions.

Conclusions: The gamified intervention seemed to be able to increase cognitive engagement and the combination of cognitive and affective engagement but not behavioral and affective engagement alone. However, positive emotions seem to play an important role in mediating the effect of gamification on engagement. In conclusion, we cannot say that gamification ”works” but that the design of an intervention, in this case, gamification, can have an impact on how participants experience the intervention.

(J Med Internet Res 2018;20(7):e247) doi:10.2196/jmir.9923

KEYWORDS

gamification; well-being; engagement; electronic mental health; mental health

Introduction

Background

Web-based interventions, in which people can improve their health from home, with or without the help of a health care professional, are increasingly used in many health care areas [1-3]. Advantages of Web-based interventions compared with face-to-face interventions are, among others, that Web-based interventions can reduce the costs of providing interventions, increase access to care for a large group of people, and are often perceived as more convenient by the users, given their flexibility and anonymity [4,5]. These interventions have been shown to be effective, for example, decreasing depressive symptoms, increasing well-being, and stimulating people to become more active [2,6-10]. However, not all Web-based interventions show beneficial effects and especially the effect sizes of interventions implemented outside the clinical setting with limited or no counselor involvement can be quite small [7,8,10].

The limited effectiveness of these interventions may be partly attributed to large nonadherence rates [11,12]. Many people who start using a Web-based intervention do not finish it or do not use the intervention in the prescribed way, which diminishes its effectiveness [13]. More recently, it has been posited that adherence (ie, using the intervention as intended by the developers) alone may not be enough for an intervention to be effective but that it is also necessary for participants to feel involved with an intervention or to be able to identify with the intervention [14,15]. Together, these factors may be called engagement, and it seems that a certain level of engagement is required for an intervention to be effective [16]. Research has shown that technology offers ample opportunities for enhancing engagement [12,17]. Gamification is one of these technological opportunities and is increasingly recommended and used to make interventions more engaging [18-21].

Gamification has been defined as “using game design elements in nongame contexts” [22]. These game design elements can be very specific, for example, the inclusion of “badges” or “levels” in the interface. They can also be broad, for example, including a storyline to make the goals clear and stimulate enduring play. The nongame part of the definition refers to gamification not being a full-fledged game, as opposed to serious games. The main goal of gamification is to increase participants’ engagement with the intervention. Multiple studies have shown the potential of gamification to increase adherence to and effectiveness of health interventions, for example, a mobile intervention for mental health [23] and a Web-based intervention for physical activity [24]. Nonetheless, many of these studies have methodological limitations and merely focus on adherence or effectiveness, thereby shedding limited light on whether and how gamification affected these variables [18,25]. Authors have indicated the need for more comparative studies (gamified vs nongamified versions of the same intervention) to isolate the effects of gamification [25].

A number of studies have shown that gamification, if used correctly, can increase intrinsic motivation for a certain behavior, for example, by satisfying certain psychological needs [26,27]. However, although motivation can be seen as necessary, it is not sufficient to achieve engagement [28]. Engagement may be seen as a multidimensional construct, consisting of behavioral, cognitive, and affective components [28,29]. In the case of Web-based health interventions, the behavioral component may refer to the usage of or adherence to the intervention [16]. It seems likely that gamification has an impact on the time participants spend in each session with the intervention, the number of exercises they perform, or how elaborately they complete each of the exercises. The cognitive and affective components are less well understood in the context of Web-based health interventions. According to a recent review, these components are often taken together as a subjective experience characterized by attention, interest, and affect [16]. Despite limited knowledge on what may constitute cognitive engagement in Web-based health interventions, from the student engagement context, in which cognitive engagement has been described as “students’ psychological involvement in learning” [30], the concept of involvement becomes apparent. Involvement relates to the importance of a product (eg, an intervention) to the individual [31] and is an important predictor of the effectiveness of Web-based mental health interventions [15]. Enjoyment is an important motivator for people to use games and might also be integral to the affective engagement with gamified interventions [19]. Enjoyment is also closely related to the concept of intrinsic motivation as a part of the self-determination theory [32]. This theory has been used to explain the appeal of games [33]. Positive emotions may also be a part of affective engagement. Because positive emotions lead to a broadening of one’s attention [34], these emotions may be especially relevant to achieve in a Web-based health intervention, wherein users are taught new ways to deal with situations. Lastly, the concept of flow is a concept related to engagement. Flow is defined as “a mental state of operation, in which a person performing an activity is fully immersed in a feeling of energized focus, full involvement, and enjoyment in the process of the activity” and is often used as a way to explain gaming behaviors [35,36]. Because it incorporates both feelings of involvement and enjoyment, it may be a state wherein both cognitive and affective engagement come together.

According to Nicholson [20], gamification can only be beneficial if it provides a “positive and meaningful game-based experience” to its users, leading to a long-term engagement. This experience seems closely related to the cognitive and affective components of engagement. It implies that gamification directly impacts the experience of the user while he or she uses the intervention. This impact should already be made during the first use of an intervention. However, as most studies focus on the longer term effects of gamification (eg, effectiveness or adherence at the end of the intervention period), little is known about the direct impact of gamification on engagement or how systems should be designed to foster this direct impact.

Objective

The goal of this study was to explore the direct impact of gamification on behavioral, cognitive, and affective engagement in the context of a Web-based mental health intervention and to explore whether and how the different components of engagement are related. To achieve this goal, an exploratory randomized experiment was carried out where participants used

a gamified or nongamified version of the same Web-based positive psychology intervention in a single session. In terms of content, both versions of the intervention were identical.

The intervention used in this experiment seeks to improve well-being. Well-being is important to achieve and maintain a healthy life and prevent mental illnesses and generally serves as a basis for resilience [37-40]. Research has shown that well-being can be improved through training and the specific intervention used in this study has also been proven effective in improving well-being. Following the positive effects of the intervention as email guided bibliotherapy [41], a Web-based version was created [42]. Although this Web-based intervention offered the opportunity to enhance the scalability of the intervention against limited costs, specific attention should be paid to engaging participants. Hence, this intervention was deemed as an ideal candidate for gamification.

Methods

Design

A between-groups experimental design was used. For the study, 2 versions of the same intervention were created (ie, a gamified version and a nongamified version). Although both versions contained the same information and exercises (ie, same texts), the information and exercises were presented in a different manner. A pilot experiment was performed in a lab setting to check the procedure and the versions of the interventions before the actual experiment was carried out in a more real-life setting. The pilot study focused on investigating the experimental procedures, not the intervention, which was pilot-tested before [42]. Our aim was to test whether participants could use the intervention without any guidance and foreknowledge in a meaningful way in one session. Therefore, the experimenters were nearby while participants used the intervention, after which they were briefly asked about their experiences. However, this formal setting seemed to influence not only the type of participants (ie, the pilot attracted mainly students who were already interested in positive psychology) but also the way they used the intervention (ie, the pilot participants used the intervention in a very focused setting without any distractions). Because this is not how the intervention will be used in real life, we decided on a different setting for the real-life experiment.

Recruitment and Participants

The study population consisted of people aged 18 years or older. Exclusion criteria were insufficient proficiency in the Dutch language (reading and writing) and the inability or unwillingness to provide informed consent. Because the University of Twente has a large proportion of German students who have learned Dutch for their studies, people of Dutch and German nationalities were able to participate as long as they had sufficient proficiency in the Dutch language. Recruitment for the pilot experiment was done through the University of Twente research participants system. Bachelor Psychology students need to earn “participant points” by participating in research studies. Overall, 19 psychology students from the University of Twente participated in the pilot experiment, of which 11 were randomized to the gamified intervention and 8 were randomized to the nongamified intervention (Table 1). Analyses showed a significant difference in nationality between the conditions with more Dutch than German participants receiving the gamified intervention (χ^2_1=8.7; *P*=.003).

For the real-life experiment, participants were recruited through the University of Twente research participation system and through convenience sampling by undergraduate psychology students who assisted in conducting the experiment as part of a Bachelor research project. Overall, 76 participants were included in the study and randomly allocated to receive either the gamified intervention (n=39) or the nongamified intervention (n=37).

One respondent in the nongamified condition had to be excluded from the analysis owing to an issue with the account the participant used for the experiment (the account had been used before so the participant was not able to complete the experimental procedure). Table 1 provides an overview of the participants included in the analyses. There were no significant differences in the demographic characteristics between the conditions.

Additional analyses showed that participants in the real-life experiment were older (22.8 vs 19.6; $F_{1,92}$=10.053; *P*=.002) and more often German (77%, 58/75 vs 37%, 7/19; χ^2_1=11.7; *P*=.001) compared with those included in the pilot experiment. There was no significant difference in gender between the pilot and real-life experiments (χ^2_1=1.3; *P*=.26).

Power Analysis

G*Power 3.1.9 (Heinrich Heine Universität, Düsseldorf, Germany) was used to calculate the required sample size for detecting a medium effect (Cohen *d*=0.5) in an independent samples *t* test (two-tailed). With 80% power at an alpha level of .05, a total sample size of 128 participants (64 per group) was needed to test the hypotheses. Unfortunately, recruitment turned out to be difficult and we did not manage to reach our intended number of participants.

Intervention

The intervention used for this experiment was called “This is your life,” a Web-based positive psychology intervention which aims to improve well-being in the general population [43] and has been proven effective as a self-help book with email counseling [41]. The Web-based gamified intervention was developed using a human-centered design [34]. Potential users that participated in the codesign process indicated the potential value of gamification and cooperated in designing the specific gamification features. Following their recommendations, we decided that the main storyline would be a user on a journey toward a flourishing life, guided by a professor. The intervention consisted of an introduction and 8 lessons that could be completed in 12 weeks. Each lesson consisted of psychoeducation and approximately 5 exercises that could be completed multiple times. In each lesson, there were approximately 2 key challenges; these were the exercises that needed to be completed to be able to continue to the next lesson.

Table 1. Participants' demographic characteristics.

Demographics	Gamified pilot (n=11) and real-life (n=39)	Nongamified pilot (n=8) and real-life (n=36)	Total pilot (n=19) and real-life (N=75)	Statistics		
				F value	χ^2_1	*P* value
Age (years), mean (SD)						
Pilot	19.6 (1,7)	20.0 (1.3)	19.7 (1.5)	0.399[a]	—	.54
Real-life	23.4 (5.6)	22.2 (1.5)	22.8 (4.2)	1.610[b]	—	.21
Sex (female), n (%)						
Pilot	8 (73)	7 (88)	15 (79)	—	0.6	.44
Real-life	28 (72)	21 (58)	49 (65)	—	1.5	.22
Nationality (Dutch), n (%)						
Pilot	10 (91)	2 (25)	12 (63)	—	8.7	.003
Real-life	10 (26)	7 (19)	17 (23)	—	0.4	.52

[a] $F_{1,17}$

[b] $F_{1,73}$

The intervention was completely self-guided; there was no guidance or feedback from a human counselor. However, the intervention itself did provide tailored feedback when a user finished a lesson and provided general feedback about how to best perform exercises at various points during each lesson. For the experiment, participants were asked to complete the introduction and 2 exercises from the first lesson in one session. These 2 exercises were "Three good things" (relive and write about 3 good things that happened today) and "Write about positive experiences" (relive and write about a beautiful memory from one's own life).

Gamified and Nongamified Version

As stated earlier, both versions of the intervention contain the same information and exercises. Differences were only in lay-out and in wording of feedback, as indicated in the next section.

Lay-Out of the Intervention Overview

In the gamified version, the overview was visualized as a map, in which the participants travel to various destinations (the different lessons). In the nongamified version, a list of lessons was provided. In both versions, the lessons that could not yet be accessed were grayed out, as seen in Figure 1.

Lay-Out of the Lesson Screen

The basic features of the lesson screen were the same in both versions (list of exercises on the left and explanation and filling out opportunity on the right, as seen in Figure 2). The gamified version showed an additional progress bar, in which the activities of the lesson were visualized; each time a mandatory activity was completed, a part of the progress bar was colored in. After finishing all the mandatory activities, participants in the gamified condition were granted a key with which they could enter the next destination. Participants in the nongamified condition were provided with a link to start the next lesson after completing the mandatory activities.

Professor and Participant Avatar

In the gamified version of the intervention, participants were guided through the intervention by an avatar of "Professor Happiness," as seen in Figures 1 and 2. Instructions and feedback appeared as a pop-up coming from the avatar. In the nongamified version, the same instructions and feedback were given through a pop-up of the info-button. The wording used in both versions was slightly adapted to appear to come from the "Professor" (eg, using "I") or from "info" (eg, using the passive form). In the gamified version, there was also room for a participant avatar (or photo), but this feature was not used in the experiments.

Badges

Participants in the gamified version earned a badge after completing the introduction and each of the lessons. These badges were shown on the right side of the screen, as seen in Figures 1 and 2. When "mousing over" these badges, a quote matching the badge's lesson was shown. Because participants in the experiments only needed to complete the introduction and some exercises in the first lesson, participants typically only saw the badge which was awarded following the completion of the introduction. The quote for this first badge was "It is good to have an end to journey toward; but it is the journey that matters, in the end–Ernest Hemingway."

Randomization and Blinding

In both the pilot and real-life experiments, participants were randomly assigned to the gamified or nongamified intervention. A random number list was created (using random.org) and participants were allocated according to this list, in the order in which they registered for the study. Randomization was double blind; the experimenters did not know to which condition the participants were assigned and the participants did not know that different versions of the intervention existed.

Figure 1. Overview of the intervention in the gamified version (left) and nongamified version (right). Source: University of Twente, Centre for eHealth and Wellbeing Research.

Figure 2. Lesson screen in the gamified version (left) and nongamified version (right). Source: University of Twente, Centre for eHealth and Wellbeing Research.

Procedure

For the pilot experiment, participants were asked to come into the lab and were seated in a cubicle, where a personal computer was set up with both the log-on page to the intervention and the questionnaire opened in a separate tab of the Internet Explorer browser. Participants gave informed consent before receiving the log-in details and instructions. The experimenter was not in the room when the participants performed the experiment but was available in case the participants experienced any (technical) issues.

In the real-life experiment, participants were made to meet face-to-face (in their home or at the University) or online to participate in the experiment. Participants gave informed consent before receiving the log-in details and the explanation of the procedure. Participants could complete the experiment at their leisure but were instructed to follow the procedure given to them.

In both the experiments, participants were asked to log into the intervention using the account details provided and to complete four tasks within the Web-based intervention in one session, regardless of condition. The four tasks were as follows: complete the tutorial, read the information of the first module, complete the exercise “Three good things” once, and complete the exercise “Write about positive experiences” once. These were the tasks we expected new users to complete during their first session. After completing these tasks, participants were free to further explore the intervention or to end the session. Completing the aforementioned tasks typically took approximately 30 minutes. After ending the session, participants were asked to fill out a Web-based questionnaire. Log data were used to assure that all participants conformed to the procedure (using the intervention for only one session and filling out the questionnaire after ending the session). Prior to the study, ethical approval was obtained from the Faculty of Behavioral Sciences Ethics Committee at the University of Twente.

Measures

Behavioral engagement was assessed by means of usage measures (ie, time spent on the intervention, the number of exercises completed, and the number of words used) gathered

through system logs (log data). For the calculation of the time spent on the intervention, "log-in" time was considered as the start time and the time when the last action was performed was considered as the end time. For this last action, it was decided that the "log-out" time would not be considered because many participants never logged out and many logged out much later (eg, 20 minutes) after performing the last action, indicating that they were not actively using the system at that time anymore, for example, completing the Web-based questionnaire. For one participant, the number of words used could not be retrieved because of a technical error.

After completing the tasks within the intervention, participants filled out a Web-based questionnaire measuring cognitive and affective engagement. For cognitive engagement, involvement was measured with the short version of the Personal Involvement Inventory (10 items, mean score 1-7, higher score means more involvement [31]). For affective engagement, positive emotions were measured with the corresponding items of the Positive And Negative Affect Schedule (PANAS, 10 items, total score 5-50, higher score means more positive emotions [44]). Enjoyment was measured with the enjoyment subscale of the Intrinsic Motivation Inventory (IMI, 7 items, mean score 1-7, higher mean score means more enjoyment [45]). Lastly, to measure both cognitive and affective engagement, flow was measured with the newly established Flow State Questionnaire of the Positive Psychology Lab (PPL-FSQ; 20 items, mean score 1-5, higher mean score means more flow [46]).

Furthermore, overall satisfaction with the intervention was assessed by asking the participants to grade the intervention from 1 to 10; higher scores indicate greater satisfaction with the intervention. Finally, in the real-life experiment, but not in the pilot test, usability was measured using the System Usability Scale (SUS, 10 items, total score 0-100, higher score means higher usability [47]). For all measures, Cronbach alpha values varied between.83 and.94 based on the data of both studies.

Analyses

Statistical analyses were performed using SPSS 23 (IBM, USA). All tests were two-tailed and the value for alpha was .05. Differences between conditions with regard to the outcome variables were investigated using one-way analysis of variance. Effect sizes are presented as Cohen *d*. Exploration of differences between conditions on single items of questionnaires (instead of on mean or sum-scores) were done using Mann Whitney *U* tests because of the ordinal level of this data from Likert-scale questions.

Within the pilot experiment, the data showed signs of not being normally distributed. Shapiro-Wilk tests were significant for the measurement of enjoyment, flow, and the number of words and exercises. Therefore, bootstrapped 95% CI of the mean differences were calculated for all outcome measures in the pilot experiment.

To explore whether and how the components of engagement were related, exploratory simple mediation analyses were conducted. More specifically, we investigated whether the influences of gamification on involvement and flow, which seems to be an effect that needed some conscious effort, was mediated by positive emotions, which seemed to be a more direct and effortless effect. A subscale of PANAS was created with items that are expected to be influenced by the gamified design (ie, interested, enthusiastic, inspired, and attentive). Mediation analyses were performed using the PROCESS macro for SPSS [48]. For each outcome (ie, involvement and flow), a separate mediation analysis was conducted. Condition (gamified or nongamified) was entered as the predictor variable, the PANAS subscale as the mediator, and either involvement or flow as the outcome variable. To test whether the indirect effect is statistically different from zero, 10,000 bootstrap CI were generated. When the corresponding bias-corrected 95% bootstrap CI did not include zero, the indirect effect was considered significant.

Results

The results of the pilot experiment conducted in the laboratory are presented in Table 2. Overall, the participants highly valued the intervention with an average grade of 7.8 and spent approximately 30 minutes using the intervention.

Table 3 presents the results of the real-life experiment. Significant differences were found for involvement and flow, whereby the gamified condition scored higher with approximately a medium effect size ($F_{1,73}$=5.919; *P*=.02 and $F_{1,73}$=4.626; *P*=.04; respectively). The total score of positive emotions did not show significant differences. Due to the various positive emotions that are measured with PANAS, we performed exploratory analyses to investigate whether there were any significant differences in distinct emotions (ie, single items of the PANAS questionnaire). With regard to the emotions "interest" and "inspiration," it was found that the gamified condition scored significantly higher than the nongamified condition (Z=−2.239; *P*=.03 and Z=−2.454; *P*=.01; respectively). There were no significant differences in the other distinct emotions. For enjoyment, usability, intervention satisfaction, and different usage measures, no significant differences were observed.

Simple mediation analyses showed that the condition indirectly influenced both involvement and flow through its effect on certain positive emotions. As seen in Figure 3 and Table 4, participants in the gamified condition scored higher on positive emotions (a=0.345), and participants who experienced more of the positive emotions interest, enthusiasm, inspiration and attentiveness, scored higher on both involvement (b=0.547) and flow (b=0.236). For both outcomes, a bias-corrected bootstrap 95% CI for the indirect effect (involvement, ab=0.188; flow, ab=0.081) based on 10,000 bootstrap samples was completely above zero (involvement, 0.017 to 0.539, flow, 0.011 to 0.204). There was no evidence that the condition influenced involvement or flow independent of the effect on the positive emotions (involvement, c'=0.460; flow, c'=0.139).

Table 2. Pilot experiment outcome variables.

Outcome	Gamified (n=11), mean (SD)	Nongamified (n=8), mean (SD)	Total (N=19), mean (SD)	Effect size (*d*)	$F_{1,17}$	*P* value	Bootstrapped 95% CI of mean difference
Positive affect	36.00 (6.01)	34.25 (2.71)	35.26 (4.87)	0.38	0.583	.46	−5.86 to 3.15
PII[a]	4.78 (1.16)	5.38 (1.10)	5.03 (1.14)	−0.53	1.271	.28	-0.53 to 1.60
IMI-E[b]	4.81 (1.23)	5.66 (1.17)	5.17 (1.25)	−0.71	2.341	.14	−0.36 to 1.91
PPL-FSQ[c]	3.61 (0.36)	3.77 (0.45)	3.68 (0.39)	−0.39	0.748	.40	−0.18 to 0.57
Satisfaction	7.55 (1.21)	8.25 (1.58)	7.84 (1.34)	−0.50	1.213	.29	−0.64 to 1.94
Number of exercises	7.55 (7.97)	3.63 (1.19)	5.89 (6.31)	0.69	1.877	.19	−8.92 to 0.21
Number of words	261.90 (271.21)[d]	205.00 (73.59)	236.61 (204.98)	0.29	0.329	.57	−244.10 to 74.79
Time	25.18 (11.94)	32.00 (7.45)	28.05 (10.62)	−0.69	2.018	.17	−2.38 to 15.97

[a]PII: Personal Involvement Inventory.

[b]IMI-E: Intrinsic Motivation Inventory, subscale Enjoyment.

[c]PPL-FSQ: Flow State Questionnaire of the Positive Psychology Lab.

[d]Based on n=10.

Table 3. Real-life experiment outcome variables.

Outcome	Gamified (n=39), mean (SD)	Nongamified (n=36), mean (SD)	Total (N=75), mean (SD)	Effect size (*d*)	$F_{1,73}$	*P* value
Positive affect	33.67 (5.47)	31.08 (7.19)	32.43 (6.44)	0.41	3.094	.08
PII[a]	4.82 (1.06)	4.18 (1.25)	4.51 (1.19)	0.55	5.919	.02
IMI-E[b]	4.71 (1.20)	4.17 (1.41)	4.45 (1.33)	0.41	3.156	.08
PPL-FSQ[c]	3.54 (0.44)	3.32 (0.45)	3.44 (0.45)	0.49	4.626	.04
SUS[d]	43.59 (14.89)	45.76 (21.49)	44.63 (18.26)	−0.12	0.263	.61
Satisfaction	6.69 (1.70)	6.36 (1.90)	6.53 (1.80)	0.18	0.634	.43
Number of exercises	4.41 (3.22)	4.36 (3.08)	4.39 (3.13)	0.02	0.005	.95
Number of words	188.10 (112.90)	164.89 (101.87)	176.96 (107.66)	0.22	0.869	.35
Time	26.95 (13.57)	29.06 (11.65)	27.96 (12.64)	−0.17	0.516	.48

[a]PII: Personal Involvement Inventory.

[b]IMI-E: Intrinsic Motivation Inventory, subscale Enjoyment.

[c]PPL-FSQ: Flow State Questionnaire of the Positive Psychology Lab.

[d]SUS: System Usability Scale.

Figure 3. Mediation models. PA: positive affect.

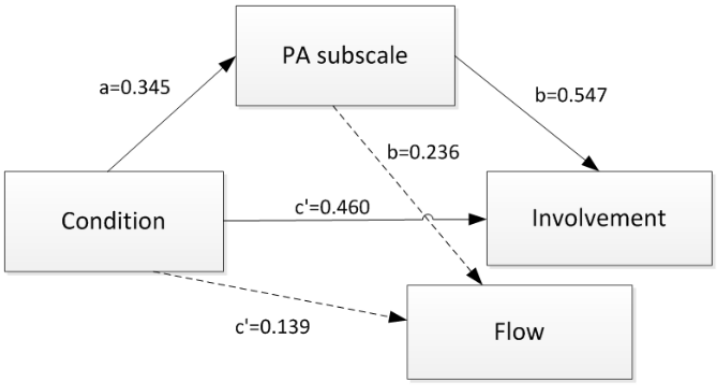


Table 4. Outcomes of the mediation models.

Variable	Coefficient	Standard error	*P* value
Positive affect subscale[a]			
Condition (a)	0.345	0.161	.04
Constant (i_1)	3.271	0.1163	<.001
Involvement[b]			
Condition (c')	0.460	0.261	.08
Positive affect subscale (b)	0.547	0.184	.004
Constant (i_2)	2.388	0.628	.003
Flow[c]			
Condition	0.139	0.099	.16
Positive affect subscale	0.236	0.070	.001
Constant	2.549	0.237	<.001

[a]R^2=.059; $F_{1,73}$=4.566, *P*=.04.

[b]R^2=.176; $F_{2,72}$=7.680, *P*=.009.

[c]R^2=.190; $F_{2,72}$=8.424, *P*=.005.

Discussion

Principal Findings

To our knowledge, this is the first study that compared the direct impact of 2 mental health interventions that have the same content and only differ in the use of gamification. This exploratory study suggests that a gamified system, in a single session, can have a positive impact on cognitive engagement by increasing the involvement participants feel with the intervention. Participants also experience more flow when working with the gamified intervention. This points toward an increase in the combination of cognitive and affective engagement. However, the gamified intervention did not seem to increase behavioral or affective engagement as such. Furthermore, gamification did not have an (negative) influence on the usability of the system. Therefore, in this study, the gamified elements did not seem to add more complexity to the system, as observed in other studies [49].

We did not see any significant differences between the conditions in the pilot experiment. Most likely, this was a power issue because there were very few participants in this experiment. However, the goal of the pilot experiment was to test the procedure of the experiment and based on the experiences of participants, changes were made in the actual study. The pilot study served its purpose.

Looking closer at the results of the real-life experiment, gamification did not seem to increase affective engagement compared to the nongamified intervention. This might be because a gamified intervention is not a game (which might be played for enjoyment and a positive experience) but only leverages some of the game design aspects. The results on the distinct positive emotions corroborate this finding: only interest and inspiration were significantly improved by gamification. A possible explanation for this finding is that these emotions relate more to the meaningfulness of the experience than to its valance. Moreover, we found that the impact of gamification on cognitive engagement was mediated by positive emotions that we expected to be influenced by the gamified design. This indicates that specific positive emotions (ie, interested, enthusiastic, inspired, and attentive) do play a role. The gamified design that was used in this study seemed to leverage these positive emotions to increase cognitive engagement with the intervention. However, because we assessed all measurements at the same time, we cannot be certain that gamification first increased these emotions, which in turn affected cognitive engagement.

Gamification did not seem to increase behavioral engagement. However, this may not be surprising because more is not always better [50]. Behavioral engagement was measured with usage data, as is common in research on Web-based interventions [16]. However, the amount of usage of a system does not always indicate the quality of usage. For instance, in this study, we asked participants to complete the activities that are likely to be sufficient for a first usage of the system. It might not be beneficial to do more exercises because you might not be able to put in the mental effort to do these extra exercises. Therefore, measuring behavioral engagement as purely the quantity of usage might not be ideal. However, because it is difficult to directly observe how participants use a Web-based intervention, especially in real life, it might not be feasible to measure the quality of behavioral engagement.

Limitations

This study has some limitations. First, this study explored the direct impact of gamification on engagement. Although we feel that this is a necessary step to start understanding the impact of gamification, we acknowledge that our findings do not address the contribution of gamification to the effectiveness of and adherence to the intervention. Longer term studies are needed for this goal. Second, in this study, we have attempted to

measure whether the gamified intervention increased cognitive and affective engagement because these aspects of the experience are posited as important for gamification to have a positive impact. However, in the literature, it is not yet defined what actually constitutes cognitive and affective engagement. Therefore, we have chosen measures that are related to these concepts and seem applicable to gamification, but there may be other measures that capture these concepts as well. Future research should investigate which measures best capture engagement. Another limitation of this study is that we only measured positive emotions after the study and not before. Lastly, this was an exploratory experiment with a relatively small sample size that did not achieve the number of participants deemed necessary based on the power analysis. Further, the participants might have had a different reason to use the intervention than the intended users of the intervention. Owing to these limitations, the results should be interpreted with caution.

Conclusions

To conclude, this study suggests that gamification, in a single session, may have a positive direct impact on involvement, flow, and the emotions "interest" and "inspiration." Thereby, the gamified intervention seemed to be able to increase cognitive engagement and the combination of cognitive and affective engagement but not behavioral and affective engagement alone. To conclude, we cannot say that gamification "works" but that the design of an intervention, in this case gamification, can have an impact on how participants experience the intervention. Although a gamified design has the potential to make Web-based mental health interventions more meaningful and relevant to its participants, it is possible that this design needs to be different for different people in different settings. Future research should investigate how to match the design of an intervention to the setting, motivation, and preferences of participants.

The fact that the design can increase cognitive engagement and impact the meaningfulness and relevance of an intervention may be especially beneficial within the context of Web-based (mental) health interventions; working on one's own well-being is important but may not necessarily need to be fun. Cognitive engagement is an important part of engagement, which is seen as an important predictor of both adherence to Web-based interventions and the effectiveness of these interventions [14,16,51]. This study is a first step in uncovering how gamification, and design in general, may enhance engagement in the context of psychological Web-based interventions and offers a starting point for creating engaging interventions.

Acknowledgments

The authors would like to thank Ellen Jakobs, Rabea Ransmann, Farina Schmelzer, and Maja Zimmermann for their assistance in conducting the study.

Conflicts of Interest

None declared.

References

1. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
2. Webb T, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
3. Cuijpers P, van SA, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
4. Andersson G, Cuijpers P. Pros and cons of online cognitive-behavioural therapy. Br J Psychiatry 2008 Oct;193(4):270-271 [FREE Full text] [doi: 10.1192/bjp.bp.108.054080] [Medline: 18827286]
5. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
6. Arnberg F, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
7. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions: A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
8. Bolier L, Haverman M, Kramer J, Westerhof GJ, Riper H, Walburg JA, et al. An Internet-based intervention to promote mental fitness for mildly depressed adults: randomized controlled trial. J Med Internet Res 2013 Sep 16;15(9):e200 [FREE Full text] [doi: 10.2196/jmir.2603] [Medline: 24041479]
9. Kelders SM, Bohlmeijer ET, Pots WTM, van GJEWC. Comparing human and automated support for depression: Fractional factorial randomized controlled trial. Behav Res Ther 2015 Sep;72:72-80. [doi: 10.1016/j.brat.2015.06.014] [Medline: 26196078]

10. Spijkerman M, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
11. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
12. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
13. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
14. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
15. Kelders S. Involvement as a Working Mechanism for Persuasive Technology, in Persuasive Technology. Springer International Publishing 2015:3-14.
16. 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]
17. Ludden G, van RTJL, Kelders SM, van GPJEWC. How to Increase Reach and Adherence of Web-Based Interventions: A Design Research Viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
18. Hamari J, Koivisto J, Sarsa H. Does gamification work? - A literature review of empirical studies on gamification. : IEEE; 2014 Presented at: 47th Hawaii international conference on system sciences (HICSS); 2014; Waikoloa, HI, USA. [doi: 10.1109/HICSS.2014.377]
19. Waite-Jones JM, Majeed-Ariss R, Smith J, Stones SR, Van RV, Swallow V. Young People's, Parents', and Professionals' Views on Required Components of Mobile Apps to Support Self-Management of Juvenile Arthritis: Qualitative Study. JMIR Mhealth Uhealth 2018 Jan 19;6(1):e25 [FREE Full text] [doi: 10.2196/mhealth.9179] [Medline: 29351898]
20. Pope L, Garnett B, Dibble M. Engaging Adolescents to Inform the Development of a Mobile Gaming App to Incentivize Physical Activity. JMIR Res Protoc 2017 Aug 29;6(8):e161 [FREE Full text] [doi: 10.2196/resprot.8113] [Medline: 28851676]
21. Leinonen A, Pyky R, Ahola R, Kangas M, Siirtola P, Luoto T, et al. Feasibility of Gamified Mobile Service Aimed at Physical Activation in Young Men: Population-Based Randomized Controlled Study (MOPO). JMIR Mhealth Uhealth 2017 Oct 10;5(10):e146 [FREE Full text] [doi: 10.2196/mhealth.6675] [Medline: 29017991]
22. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining gamification. : ACM; 2011 Presented at: Proceedings of the 15th International Academic MindTrek; 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
23. Ahtinen A, Mattila E, Välkkynen P, Kaipainen K, Vanhala T, Ermes M, et al. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR Mhealth Uhealth 2013 Jul 10;1(2):e11 [FREE Full text] [doi: 10.2196/mhealth.2596] [Medline: 25100683]
24. Thorsteinsen K, Vittersø J, Svendsen GB. Increasing physical activity efficiently: an experimental pilot study of a website and mobile phone intervention. Int J Telemed Appl 2014;2014:746232 [FREE Full text] [doi: 10.1155/2014/746232] [Medline: 24963290]
25. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interventions 2016 Nov;6:89-106. [doi: 10.1016/j.invent.2016.10.002]
26. Mekler E, Brühlmann F, Tuch A, Opwis K. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Computers in Human Behavior 2017 Jun;71:525-534. [doi: 10.1016/j.chb.2015.08.048]
27. Sailer M, Hense J, Mayr S, Mandl H. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior 2017 Apr;69:371-380. [doi: 10.1016/j.chb.2016.12.033]
28. Appleton J, Christenson S, Furlong M. Student engagement with school: Critical conceptual and methodological issues of the construct. Psychology in the Schools 2008;45(5):369-386. [doi: 10.1002/pits.20303]
29. Graffigna G. Is a Transdisciplinary Theory of Engagement in Organized Settings Possible? A Concept Analysis of the Literature on Employee Engagement, Consumer Engagement and Patient Engagement. Front Psychol 2017;8:872 [FREE Full text] [doi: 10.3389/fpsyg.2017.00872] [Medline: 28729843]
30. Archambault I, Janosz M, Morizot J, Pagani L. Adolescent behavioral, affective, and cognitive engagement in school: relationship to dropout. J Sch Health 2009 Sep;79(9):408-415. [doi: 10.1111/j.1746-1561.2009.00428.x] [Medline: 19691715]
31. Zaichkowsky J. Measuring the involvement construct. Journal of Consumer Research 1985;12(3):341-352.

32. 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]
33. Rigby S, Ryan R. Glued to games: How video games draw us in and hold us spellbound. Santa Barbara, California, United States: ABC-CLIO; 2011:186.
34. Fredrickson BL. The broaden-and-build theory of positive emotions. Philos Trans R Soc Lond B Biol Sci 2004 Sep 29;359(1449):1367-1378 [FREE Full text] [doi: 10.1098/rstb.2004.1512] [Medline: 15347528]
35. Csikszentmihalyi M. Flow and the Psychology of Discovery and Invention. New York: HarperPerennial; 1997:39.
36. Jegers K. Pervasive game flow: understanding player enjoyment in pervasive gaming. Computers in Entertainment (CIE) 2007;5(1):9. [doi: 10.1145/1236224.1236238]
37. Keyes CLM, Dhingra SS, Simoes EJ. Change in level of positive mental health as a predictor of future risk of mental illness. Am J Public Health 2010 Dec;100(12):2366-2371. [doi: 10.2105/AJPH.2010.192245] [Medline: 20966364]
38. Lamers S, Westerhof G, Glas C, Bohlmeijer E. The bidirectional relation between positive mental health and psychopathology in a longitudinal representative panel study. The Journal of Positive Psychology 2015 May 27;10(6):553-560. [doi: 10.1080/17439760.2015.1015156]
39. Duckworth AL, Steen TA, Seligman MEP. Positive psychology in clinical practice. Annu Rev Clin Psychol 2005;1:629-651. [doi: 10.1146/annurev.clinpsy.1.102803.144154] [Medline: 17716102]
40. Ryff C. Psychological well-being revisited: advances in the science and practice of eudaimonia. Psychother Psychosom 2014;83(1):10-28 [FREE Full text] [doi: 10.1159/000353263] [Medline: 24281296]
41. Schotanus-Dijkstra M, Drossaert CHC, Pieterse ME, Boon B, Walburg JA, Bohlmeijer ET. An early intervention to promote well-being and flourishing and reduce anxiety and depression: A randomized controlled trial. Internet Interventions 2017;9:15-24. [doi: 10.1016/j.invent.2017.04.002]
42. Ludden G, Kelders S, Snippert B. This is your life!. In: Spagnolli A, Chittaro L, Gamberini L, editors. PERSUASIVE 2014. Lecture Notes in Computer Science, vol 8462. Cham: Springer; 2014.
43. Bohlmeijer E, Hulsbergen M. Dit is jouw leven. Ervaar de effecten van de positieve psychologie (This is Your Life. Experience the Effects of Positive Psychology). Amsterdam: Uitgeverij Boom; 2013:188.
44. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [Medline: 3397865]
45. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989 Mar;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
46. Magyaródi T, Nagy H, Soltész P, Mózes T, Oláh A. Psychometric properties of a newly established flow state questionnaire. The Journal of Happiness & Well-Being 2013;1(2):85-96. [doi: 10.1556/Pszicho.33.2013.1.2]
47. Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry 1996;189(194):4-7 [FREE Full text]
48. Hayes A. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. In: Guilford Publications. New York City: Guilford Publications; 2017.
49. Boendermaker WJ, Prins PJM, Wiers RW. Cognitive Bias Modification for adolescents with substance use problems--Can serious games help? J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):13-20 [FREE Full text] [doi: 10.1016/j.jbtep.2015.03.008] [Medline: 25843611]
50. Sieverink F, Kelders SM, van Gemert-Pijnen JEWC. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
51. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]

Abbreviations

IMI: Intrinsic Motivation Inventory
PANAS: Positive and Negative Affect Schedule
PII: Personal Involvement Inventory
PPL-FSQ: Flow State Questionnaire of the Positive Psychology Lab
SUS: System Usability Scale

Edited by G Eysenbach; submitted 24.01.18; peer-reviewed by E Mekler, K Blondon, P Lindner; comments to author 25.03.18; revised version received 18.05.18; accepted 29.05.18; published 26.07.18.

Please cite as:
Kelders SM, Sommers-Spijkerman M, Goldberg J
Investigating the Direct Impact of a Gamified Versus Nongamified Well-Being Intervention: An Exploratory Experiment
J Med Internet Res 2018;20(7):e247
URL: http://www.jmir.org/2018/7/e247/
doi:10.2196/jmir.9923
PMID:30049669

Original Paper

How Online Communities of People With Long-Term Conditions Function and Evolve: Network Analysis of the Structure and Dynamics of the Asthma UK and British Lung Foundation Online Communities

Sagar Joglekar[1], MS; Nishanth Sastry[1], PhD; Neil S Coulson[2], PhD; Stephanie JC Taylor[3], MD; Anita Patel[3], PhD; Robbie Duschinsky[4], PhD; Amrutha Anand[5], MPH; Matt Jameson Evans[5], MB BS; Chris J Griffiths[3], D Phil; Aziz Sheikh[6], MD; Pietro Panzarasa[7], PhD; Anna De Simoni[3], PhD

[1]Department of Informatics, King's College London, London, United Kingdom
[2]School of Medicine, University of Nottingham, Nottingham, United Kingdom
[3]Asthma UK Centre for Applied Research, Barts Institute of Population Health Sciences, Queen Mary University of London, London, United Kingdom
[4]Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom
[5]HealthUnlocked, London, United Kingdom
[6]Asthma UK Centre for Applied Research, Usher Institute of Population Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom
[7]School of Business and Management, Queen Mary University of London, London, United Kingdom

Corresponding Author:
Anna De Simoni, PhD
Asthma UK Centre for Applied Research
Barts Institute of Population Health Sciences
Queen Mary University of London
58 Turner Street
London, E1 2AB
United Kingdom
Phone: 44 207 882 2520
Fax: 44 20 78822
Email: a.desimoni@qmul.ac.uk

Abstract

Background: Self-management support can improve health and reduce health care utilization by people with long-term conditions. Online communities for people with long-term conditions have the potential to influence health, usage of health care resources, and facilitate illness self-management. Only recently, however, has evidence been reported on how such communities function and evolve, and how they support self-management of long-term conditions in practice.

Objective: The aim of this study is to gain a better understanding of the mechanisms underlying online self-management support systems by analyzing the structure and dynamics of the networks connecting users who write posts over time.

Methods: We conducted a longitudinal network analysis of anonymized data from 2 patients' online communities from the United Kingdom: the Asthma UK and the British Lung Foundation (BLF) communities in 2006-2016 and 2012-2016, respectively.

Results: The number of users and activity grew steadily over time, reaching 3345 users and 32,780 posts in the Asthma UK community, and 19,837 users and 875,151 posts in the BLF community. People who wrote posts in the Asthma UK forum tended to write at an interval of 1-20 days and six months, while those in the BLF community wrote at an interval of two days. In both communities, most pairs of users could reach one another either directly or indirectly through other users. Those who wrote a disproportionally large number of posts (the superusers) represented 1% of the overall population of both Asthma UK and BLF communities and accounted for 32% and 49% of the posts, respectively. Sensitivity analysis showed that the removal of superusers would cause the communities to collapse. Thus, interactions were held together by very few superusers, who posted frequently and regularly, 65% of them at least every 1.7 days in the BLF community and 70% every 3.1 days in the Asthma UK community. Their posting activity indirectly facilitated tie formation between other users. Superusers were a constantly available resource, with a mean of 80 and 20 superusers active at any one time in the BLF and Asthma UK communities, respectively. Over time,

the more active users became, the more likely they were to reply to other users' posts rather than to write new ones, shifting from a help-seeking to a help-giving role. This might suggest that superusers were more likely to provide than to seek advice.

Conclusions: In this study, we uncover key structural properties related to the way users interact and sustain online health communities. Superusers' engagement plays a fundamental sustaining role and deserves research attention. Further studies are needed to explore network determinants of the effectiveness of online engagement concerning health-related outcomes. In resource-constrained health care systems, scaling up online communities may offer a potentially accessible, wide-reaching and cost-effective intervention facilitating greater levels of self-management.

(J Med Internet Res 2018;20(7):e238) doi:10.2196/jmir.9952

KEYWORDS

asthma; chronic obstructive pulmonary disease; COPD; network analysis; online community; online forums; superusers; self-management; digital health social network

Introduction

Background

Online communities have the potential to influence health and health care. Recent studies have suggested that the participation of people with long-term conditions (LTCs) in online communities (1) improves illness self-management [1], (2) produces positive health-related outcomes [2-4], (3) facilitates shared decision-making with health care professionals [5,6], and (4) may even reduce mortality [7].

There is also evidence that self-management support interventions can reduce health service utilization [8,9].

Online communities have experienced an upsurge in popularity among people with chronic respiratory conditions such as cystic fibrosis [10], asthma [11], pulmonary hypertension [12] and chronic obstructive pulmonary disease (COPD) [13]. More than 15 million people in England suffer from a long-term condition or disability, and they account for at least 50 percent of all general practitioner appointments [14,15]. Thus, assessing how these online communities function and evolve can have important implications for health care provision.

This form of "user-led self-management" of LTCs bears similarities with the "expert patient" model, an approach to self-management of LTCs produced by the United Kingdom (UK) Department of Health in 2001 [16]. Evidence of the effectiveness of conventional off-line self-management programs based on the expert patient model, though, has been weak [17]. Clinic-based self-management programs often failed because of: (1) lack of awareness and engagement among patients and staff, (2) failure to consider low health literacy or cultural norms, (3) lack of attention to the need for family and social support, and (4) a fragmented approach to the provision of health and social care [18]. Although online health communities can be seen as an extension of the expert patient model, network effects, in addition to the online disinhibition effect [19], make them a distinct and unique complex intervention mechanism.

On average, one in four people with an LTC who use the Internet tries to engage online with others with similar health-related concerns [20]. In particular, it has been suggested that the value of participating in an online community lies in the possibility of gaining access to a range of people and resources quickly, easily [21], and anonymously [4], as well as obtaining tailored information and emotional support [1,22-26]. However, most of this evidence comes from qualitative studies [1,27], whereas only recent years have witnessed an increasing interest in quantitative assessments of online communities as intervention mechanisms [28-33]. Recent studies have been concerned with the users' unequal contributions and engagement patterns, and with the role of superusers. However, the contribution of superusers to the sustainability of online health communities and their structural properties remains mostly unclear.

The potential future integration of online health support systems with formal health care provision should be underpinned by a better understanding of how they are used and by evidence of their effectiveness. Indeed, as suggested by the Medical Research Council [34], integrating online support systems with the more traditional health care provision would require the identification and comparative assessment of potential alternative intervention mechanisms.

An expanding body of literature concerned with social network analysis has examined the structural patterns of relations among interacting actors and the social mechanisms that enable them to gain access to valuable resources [35]. There is also increasing evidence that network approaches can be applied to understanding the users' "expertise" [36], their interactions, and network effects on health-related outcomes in online health communities [37,38]. Uncovering the mechanisms underlying the formation of successful social networks requires a study of how online connections among people, namely the social ties or links, emerge and evolve, and how groups of individuals gradually grow in membership and become interconnected with one another. These processes of tie creation and group formation in online patients' communities are still mostly unexplored [1].

In this study, we performed a network analysis of the structure and dynamics of two online communities of people with LTCs. We chose the Asthma UK and the British Lung Foundation (BLF) communities as an exemplar of such communities because their users typically suffer from chronic respiratory conditions. In particular, while Asthma UK users typically suffer from a respiratory condition characterized by variable and recurring symptoms, BLF users represent a more heterogeneous population of participants affected by different diseases linked to chronic symptoms of breathlessness (eg, COPD, pulmonary fibrosis, cystic fibrosis, and lung cancer).

Textbox 1. Research questions.

1. What is the network structure of online communities for people with long-term conditions, and how do they function and evolve over time?
2. Does posting activity follow a time pattern?
3. Are there (a minority of) users with a special role in maintaining integration and cohesion of the community?
4. Do superusers write their posts uniformly over time or do they produce peaks of activity separated by periods of inactivity?
5. For how long do superusers remain active in an online community?
6. Are superusers help-seekers or help-givers?
7. Do superusers preferentially write posts to each other or to users who write relatively few posts?
8. Is there any association between users' interaction patterns and their potential for enhancing peer self-management support in the community?
9. Do online health communities function and evolve in the same way as other real-world complex systems?

We aimed to uncover and understand how these communities function and evolve, and the role that some users have in maintaining integration and cohesion (see Textbox 1 for research questions). Ultimately, this study provides evidence for gauging the effectiveness of different interaction patterns and the users' structural positions and their potential for enhancing and sustaining health online communities as scalable self-management support interventions.

Methods

Data Collection

Data were collected by HealthUnlocked [39], the online platform provider of the Asthma UK and BLF communities. Registered users can choose to either write posts publicly or send private posts to one another. In the latter case, posts are shared between 2 users only, whereas when posts are written publicly, a large number of users can become connected through threads of posts. Only posts that were shared publicly were collected and analyzed. For this study, user identifiers (IDs) were anonymized by HealthUnlocked, and no demographic information was collected. The data sets included posts and their metadata (ie, the anonymized user ID numbers), user roles (eg, user, administrator, or moderator), date of posting, the hierarchical level of the post within the corresponding thread, and the dates in which the users joined and left the community. Both communities were moderated, and HealthUnlocked moderators (identified through metadata linked to posts) were included in the analysis to assess their contribution and compare it with other users. Online communities on the HealthUnlocked platform benefit from additional functionalities compared to other online forums, such as built-in patient groups that moderate the content. In particular, the content accessed by users is tailored to their interests, and profiles highlight users' condition, chosen community, medications and treatments they use or find interesting. No data were collected on participants' characteristics, though only people declaring themselves to be older than 16 years were permitted to create an account and take part in the online communities.

Data Analysis

We looked at the number of users, the number of posts and connections per user and posting frequency. A connection (ie, a tie, link, or edge) was established from one user to another when the former replied to a post by the latter (see Textbox 2 for network analysis terminology). The pattern of connections generated over time through the cumulative number of posts and replies was examined. We were interested not just in the number of posts and responses but in who responded to whom, and when. To this end, we used social network analysis [40] to visualize and study the structure of the relationships between users. Both visualization and analysis were conducted using the Gephi software. The network analysis was carried out through additional custom computer code in python. Descriptive analysis of the networks (ie, number of users, posts, and posting frequency) were calculated using the Pandas library, an open source library providing data structures and analysis tools for the Python programming language.

As a result of the small percentage of users who wrote posts to a disproportionally high number of users, the users' activity showed long-tailed distributions. Therefore, our analysis was based not only on means and standard deviations but also on medians.

To uncover time patterns in posting activity, we used Fourier transforms of the time series of the users' activity [46], a known method used for the analysis of signals. Through Fourier transforms, we identified the frequency components, called harmonics, that together made up the posting activity stream. In other words, we regarded the posting activity over the entire observation period in both communities as a complex signal and identified the frequency components that made up such a signal. This analysis was performed using custom code in Scipy, a Python-based scientific computing library.

The "rich-club" coefficient is a metric designed to measure the extent to which well-connected users tend to connect with one another to a higher degree than expected by chance [43]. To this end, for each value k of a node's degree (ie, the number of other users a given user is connected with), we computed the ratio between the number of actual connections between nodes with degree k or larger and the total possible number of such connections [47]. We then divided this ratio by the one obtained on a corresponding random network with the same number of nodes and degree distribution (ie, the probability distribution of the degrees over the whole network) as the real network, but in which links were randomly reshuffled between nodes. Thus, the rich-club coefficients may take values lower or higher than 1, depending on whether the real network has a higher or lower

tendency to coalesce into rich clubs than randomly expected. In particular, networks that display a high rich-club coefficient (ie, greater than 1) are also said to show a "rich-club effect," namely the tendency to organise into a hierarchical structure in which highly connected nodes preferentially create tightly knit groups with one another, thus generating exclusive clubs of (topologically) rich nodes, as illustrated in previous work [48].

In our study, superusers were defined according to their cumulative activity over the entire observation period. In total, we identified 400 superusers. To uncover how many superusers were active within each week, we detected how many unique users, among the 400 identified over the entire period, were active within that time window.

Following Zhang et al [36], the "z-score" was used as a proxy for users' expertise. According to this measure, replying to many questions suggests one's expertise, while asking questions indicates lack of expertise. In our analysis, we treated anyone starting a thread as a help-seeker, and anyone commenting on the thread as a help-giver [36]. Accordingly, the proposed *z*-score aims to capture the combined help-seeking and help-giving patterns. To this end, for each user, we measured how many standard deviations the observed total number of the user's help-giving posts lies above or below the expected number of help-giving posts for the whole system. We extended the approach proposed by Zhang et al by empirically assessing the probability of posting and answering a question across all users over the entire observation period. In the BLF community, we found that the probability of answering is P_a=2/3, while the probability of posting is P_q=1/3. We assumed a Bernoulli process of posting an answer or a question to the forum, with probabilities defined as above. The *z*-score for a given user *i* was calculated according to equation (a) in Figure 1, where a_i refers to the total number of answers user *i* posted to the forum, q_i is the total number of questions user *i* asked in the forum, and $n_i=a_i+q_i$ is the total number of messages posted by user *i*.

Textbox 2. Network analysis terminology.

- Degree: the number of connections a user has established with other users through posts
- Ego(-centred) network: the subset of connections linking a focal user—"ego"—directly to other users—"alters"—and connections linking these alters with each other
- Largest component: the network component (see below) with the largest number of members.
- Network Component: a subset of the network in which all members are directly or indirectly connected with one another (ie, all pairs of nodes in the subset are reachable through at least one tie) [41,42]. Each isolated user can be regarded as a separate component
- Node: individual user in an online community
- Rich-club coefficient: the degree to which highly connected users preferentially connect to each other to a higher degree than would be expected by chance. In a community with a rich-club coefficient higher than 1, users who post to many others preferentially communicate with each other, thus forming rich clubs. Conversely, in a community with a rich-club coefficient lower than 1, users who post to many others preferentially communicate with those who post to few others, thus generating an anti-rich-club behavior [43]
- Root post: the initial post in a thread of posts
- Superusers: top 1% of users characterised by the largest number of posts written in the community over the entire observation period [44]
- Tie, link, edge: online connection from a user to another, created when the former writes a post to the latter
- Triad: a group of 3 users—nodes *i*, *j*, and *u* —forming a path of length 2 (ie, node *i* is connected to node *j*, and node *j* is connected to node *u*). When node *i* is also connected to node *u*, the path is closed, forming a loop of length 3 or a triangle
- *z*-score: a measure of users' expertise, capturing the users' combined "help-seeking" and "help-giving" patterns. If a user writes help-seeking and help-giving posts equally often, then the user's *z*-score would be equal to zero. Conversely, if a user writes more (or fewer) help-giving posts than help-seeking ones, then the *z*-score would be positive (or negative) [36,45]

Figure 1. The *z*-score used as a proxy for users' expertise.

(a) $Z_{score_i} = \frac{a_i - 2q_i}{\sqrt{2(a_i+q_i)}}$

(b) $\mu_{random} = n_i P_a$

(c) $\mu_{random} = n_i P_a = \frac{2n_i}{3}$

(d) $\sigma_{random} = \sqrt{n_i P_a (1 - P_a)}$

(e) $\sigma_{random} = \frac{\sqrt{2n_i}}{3}$

(f) $Z_{score_i} = \frac{a_i - \mu_{random}}{\sigma_{random}}$

(g) $Z_{score_i} = \frac{a_i - \frac{2n_i}{3}}{\frac{\sqrt{2n_i}}{3}}$

(h) $Z_{score_i} = \frac{a_i - 2q_i}{\sqrt{2(a_i+q_i)}}$

To obtain Z_{score_i}, let us define a random user that posts the same total number of messages n_{random} to the forum as user i (ie, $n_{random}=n_i$). We would expect this random user to post an average number of answers to the forum given by equation (b). Plugging in the value of $P_a=2/3$, we obtained equation (c). Similarly, we would expect the random user to post answers with a standard deviation given by equation (d). Plugging in the value of $P_a=2/3$, we obtained equation (e). To measure how many standard deviations above or below the expected random value a user i lies, we then computed Z_{score_i} according to equation (f). Plugging in the values of μ_{random} and σ_{random}, we obtained equation (g). Finally, by substituting $n_i=a_i+q_i$, we obtained equation (h).

Ethical Considerations

Permission to research was obtained from Asthma UK and the BLF before starting the study. The research protocol was examined, and permission to research was obtained from Asthma UK, BLF charities and HealthUnlocked. The study was examined by the institutional Research Ethics board at Queen Mary University of London and was exempt from full review.

Results

Descriptions of Data Sets

The data sets span, respectively, 10 years for the Asthma UK and 4 years for the BLF communities (see Table 1).

Despite the shorter time span, as a result of the larger number of users, the number of posts in the BLF community was higher than in Asthma UK, namely 875,151 compared to 32,780 respectively. Moreover, BLF users wrote a higher number of posts per user and were connected with a higher number of other users when compared with people in the Asthma UK forum (see Figure 2). In both communities, 60%-70% of registered users wrote no posts (ie, they were lurkers). Users who wrote more than one post contributed with a median of 8 (range 2-8947) and 5 (range 2-1068) posts in the BLF and Asthma UK communities, respectively.

The number of official moderators among the highly active users was negligible; there were no moderators in the top 5% contributors to BLF and only 2 in the top 5% for Asthma UK. Thus, our network analysis predominantly reflects content originated from registered users.

When classified according to posting activity (ie, number of posts written to the forum), the top 5% users contributed to a substantial proportion of all posts: 58% and 79% in the Asthma UK and BLF communities, respectively. Superusers were those who made a high number of connections with other users in both Asthma UK and BLF communities (see nodes of large size in Figure 2). Asthma UK superusers made a lower number of connections than BLF ones. The posting activity of these superusers will be analyzed in more detail in subsequent sections.

Table 1. Description of the Asthma UK and British Lung Foundation data sets.

Variables	Asthma UK	British Lung Foundation
Data set time span (mm/dd/yyyy)	02/03/2006-06/09/2016	13/04/2012-06/09/2016
Total time (weeks)	548	230
Total number of posts, n	32,780	875,151
Number of posts with reply, n (%)	28,615 (87.3)	815,184 (93.1)
Number of posts with no reply, n (%)	4165 (12.7)	59,967 (6.9)
Total number of users, n	3345	19,837
Users who wrote ≥1 post, n (%)	1053 (31.5)	7814 (39.4)
Users who wrote 1 post, n (%)	331 (31.4)	1186 (15.2)
Users who wrote >1 post, n (%)	722 (68.6)	6628 (84.8)
Registered users who never posted (ie, lurkers), n (%)	2292 (68.5)	12,023 (60.6)
Number of posts per user, mean (SD)	14.2 (55.0)	66.9 (75.1)
Number of posts per users who posted >1, median (range)	5.1 (2-1068)	8.0 (2-8947)
Number of posts per users who posted >1, mean (SD)	20.4 (65.6)	88.1 (458.6)
Posts contributed by top 1% superusers, n (%)	10,457 (31.9)	426,198 (48.7)
Number of connections per user, mean (SD)	2.1 (5.9)	17.6 (69.0)
Number of connections per user, median (SD)	1.0 (5.9)	1.0 (69.0)
Number of connections per top 1% superuser, mean (SD)	10.5 (16.5)	141.0 (174.0)
Number of connections per top 1% superuser, median (SD)	7.0 (16.5)	70.0 (174.0)

Figure 2. Cumulative networks across the time span analyzed. Each node represents a user. (A) Asthma UK users (around 1000); (B) British Lung Foundation users (around 8000). The coloring of nodes is based on modularity membership and the size of the node is proportional to its degree (ie, the number of connections with other users).

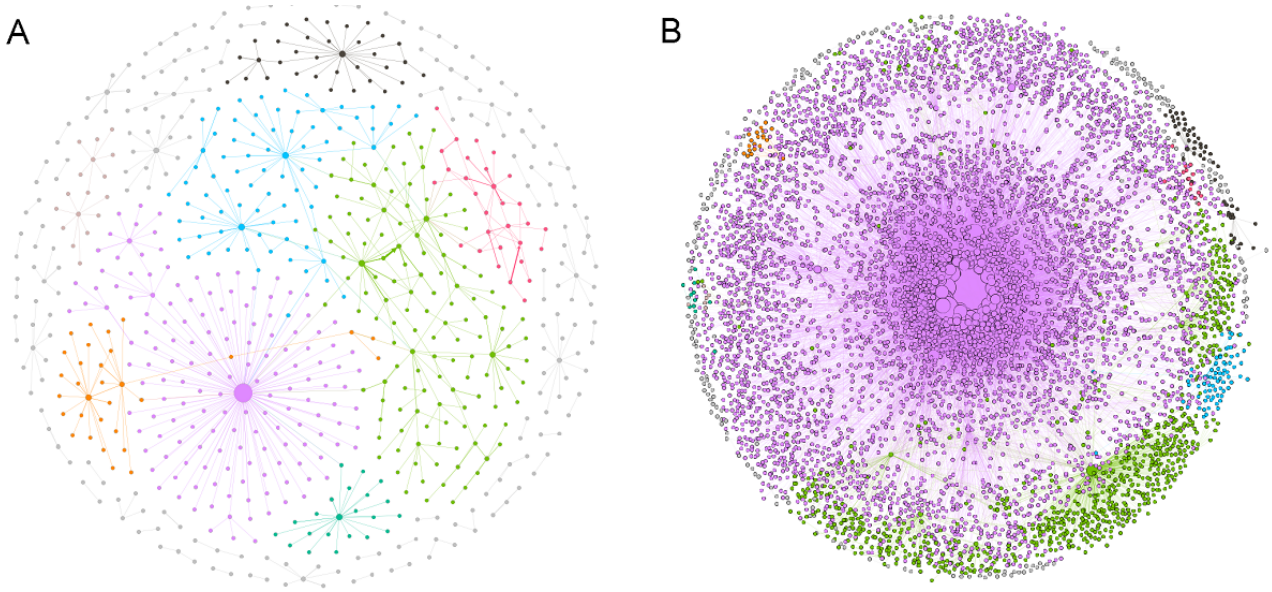
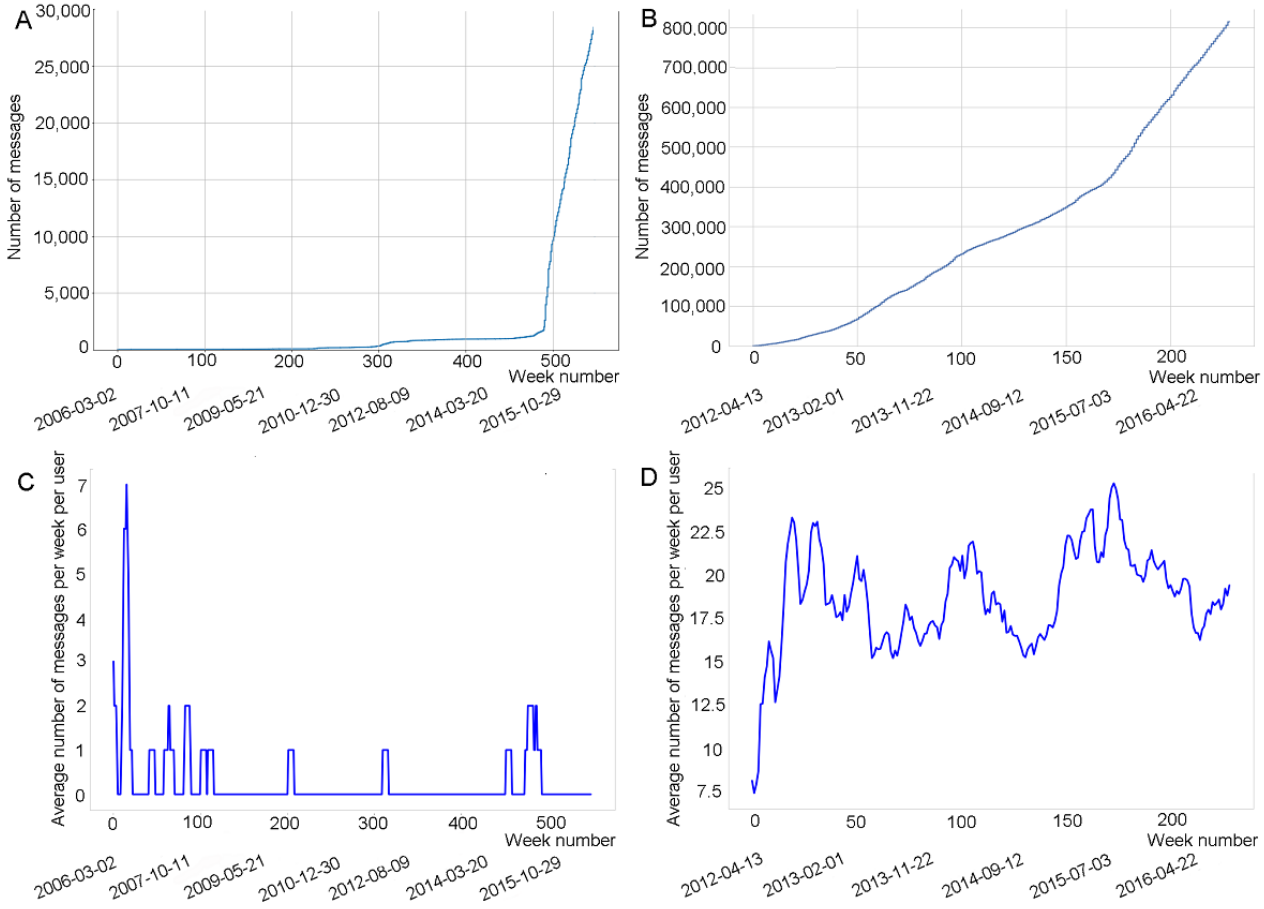


Figure 3. Cumulative distributions of the number of posts as a function of time (weeks) within the Asthma UK (A) and the British Lung Foundation (B) communities. Calendars dates are reported below week numbers. Panels C and D illustrate the average number of posts per user per week within Asthma UK and British Lung Foundation, respectively.



Posting Activity

The cumulative number of messages posted grew uniformly over time in the BLF community. By contrast, in 2015, the Asthma UK forum witnessed a substantial increase in posting activity, at a time coinciding with its move to the HealthUnlocked platform (see Figure 3A and B). This increase in activity can be attributed to the online community functionalities offered by HealthUnlocked, as described in the Methods.

The number of posts per user per week oscillated around a decreasing and an increasing trend (Figure 2C and D), while at the same time the number of posts always went up over the study period (Figure 1A and B). This suggests that there were intervals of time during which the rate of increase in new users was larger than the rate of increase in total posts. Moreover, in the Asthma UK forum users wrote according to two time patterns—they posted at an interval of 1-20 days or 6 months (Figure 4A), while those in the BLF community at an interval of 2 days (Figure 4B).

As more users joined the communities and connected to one another through online posts, distinct groups of connected users started to emerge. These groups, called network components (see Textbox 2), have fundamental implications for the effectiveness of processes of network dynamics such as information diffusion [49]. In a relatively short period, both communities underwent the formation of the "largest component" of connected users, namely a connected subset of users whose size increasingly outgrew the size of all other components (see Figures 1 and 4, and Multimedia Appendices 1 and 2). The largest connected components in both communities included 60%-100% of users.

Figure 5 suggests that, as time went by, the number of forum participants and their posting activity increased, and the proportion of users who were part of the largest components decreased. This finding was expected because the number of posts also rose exponentially, yet at times at a lower rate than the one at which new users joined the communities (see Figure 1C and D). It, therefore, became more difficult for the network to self-organize into a connected component that would include 100% of the users. Figure 5A also shows that around week 450, when the forum moved to the HealthUnlocked platform, a larger fraction of users began to join the largest connected component, thus highlighting the role that the new online platform played in strengthening the connectedness of the network (see also Figure 3A and B).

Superusers

Superusers represented a small minority (ie, 1%-5%) within both communities but were responsible for a high proportion of the posting activity and the functioning of the communities.

Superusers' Role

Sensitivity analysis showed that the removal of users with the largest number of connections caused the largest component to collapse (see Figure 6), thus suggesting that both communities and lines of communication within them were held together precisely by these highly connected users. In online communities, the existence of groups of highly connected users is critical for information diffusion [50].

Figure 4. Periodicity of posting activity in Asthma UK (A) and the British Lung Foundation (B), measured through the Fast Fourier Transform (FFT). The component frequencies are denoted by *f* and are inverted to produce time period in days.

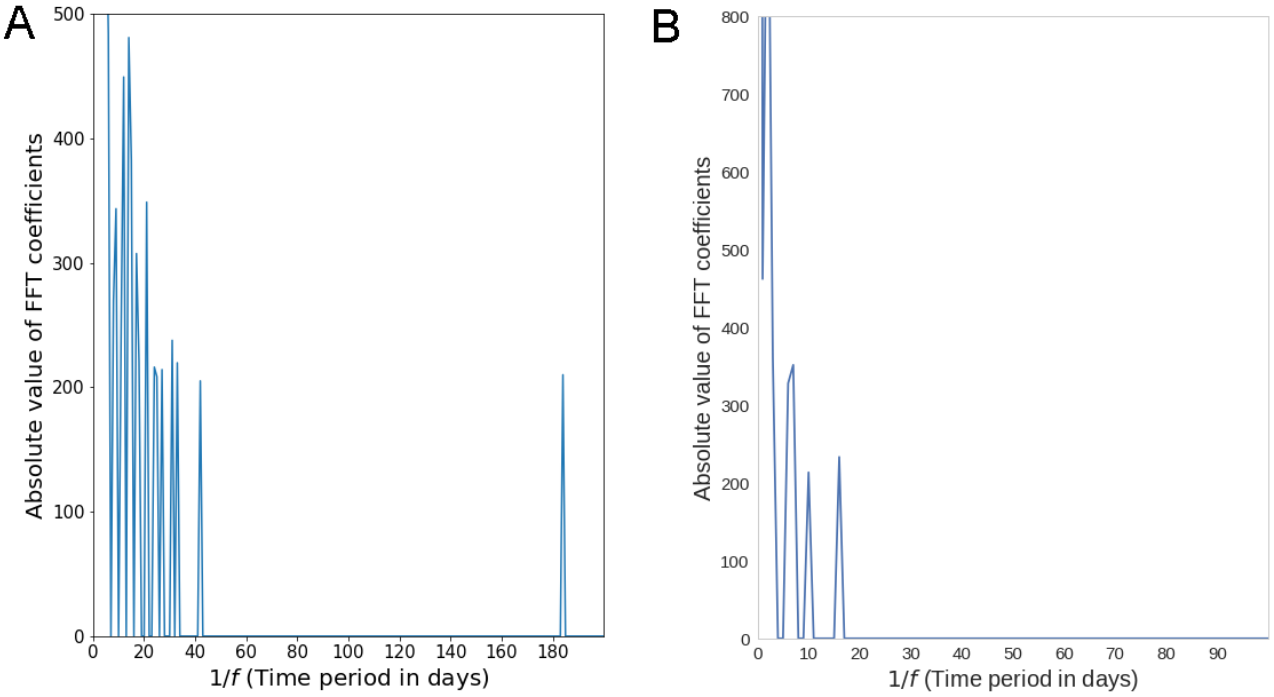


Figure 5. Fraction of users that are part of the largest component as a function of time (weeks) for Asthma UK (A) and the British Lung Foundation (B).

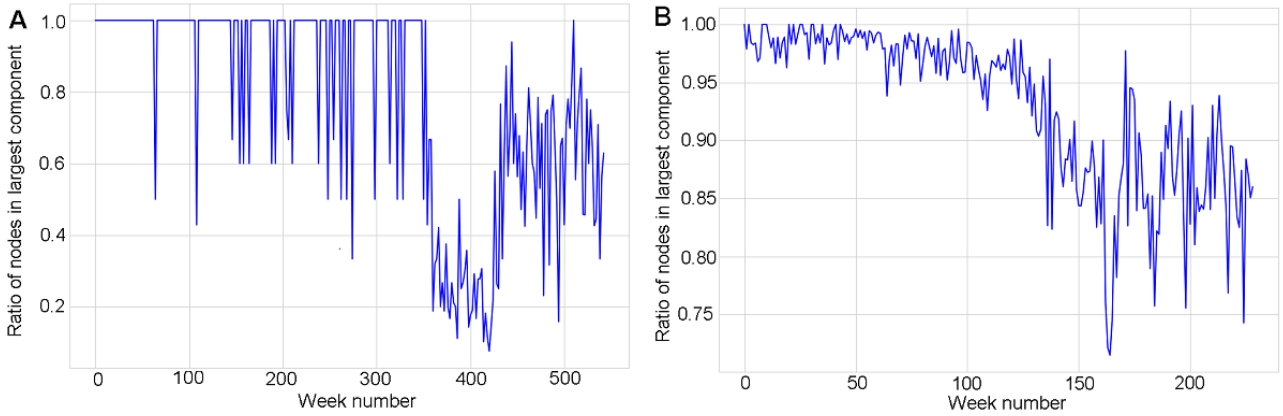


Figure 6. Sensitivity analysis: targeted removal of nodes (users) starting from the most connected ones within Asthma UK (A) and the British Lung Foundation (B).

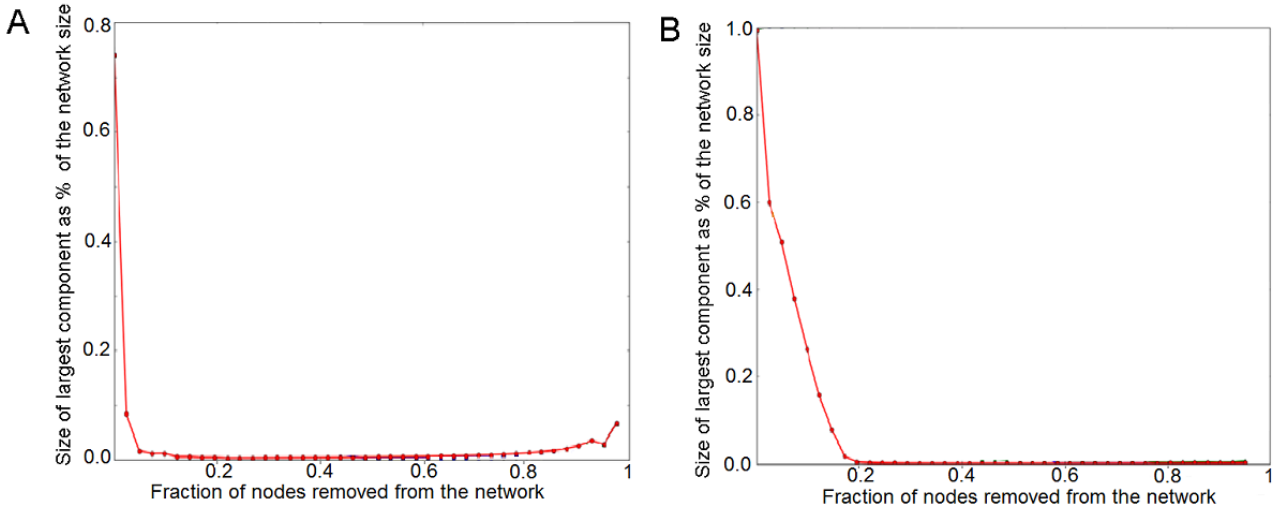


Figure 6 suggests that it only takes the removal of the top 5% users by degree of connectivity for the largest connected component to collapse to 10% and 50% of its original size in the Asthma UK and BLF communities, respectively. This corresponds to the removal of about 50 and 400 users in the 2 communities, respectively. These results shed light on how many superusers are needed to sustain discussions and to serve the needs of users in large communities of people with LTCs.

Superusers and the Rich-Club Effect

Both Asthma UK and BLF communities were characterized by a low rich-club coefficient, which was consistently lower than 1 (see Figure 7). This anti-rich-club behavior, namely the tendency to run counter to the formation of a rich club, suggests that in both communities highly connected superusers preferentially communicated with poorly connected ones or, alternatively, that superusers tended to avoid each other and instead communicated with those who were only connected with very few others.

Anti-rich-club behavior may suggest competition between superusers or merely the organization of the communities into groups of users characterized by different degrees of "expertise" or commitment: one group including the few committed experts and another including the vast majority of those seeking information when needed. It would, therefore, come as no surprise if the former were to communicate with the latter to a greater extent than randomly expected. We shall investigate this hypothesis further below.

Involvement of Superusers Over Time

We have shown that the connectedness of both communities depends crucially on the presence and activities of superusers, who committed a significant amount of their time to writing posts and targeting new users. We now look at whether their activity was concentrated in relatively short periods of time or instead it was uniformly distributed over time. How superusers' involvement is distributed over time may have fundamental implications for the cohesion of the whole system precisely in light of the role these users play.

Figure 7. Rich-club coefficient as a function of the richness parameter (ie, users' degree).

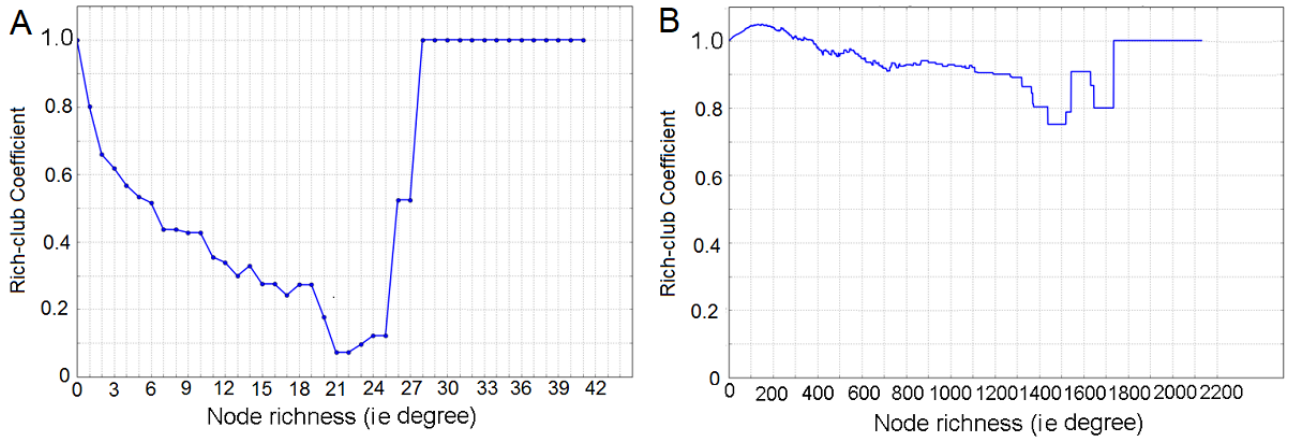


Figure 8. Number of unique users among the top 400 superusers as a function of time (weeks) within Asthma UK (A) and the British Lung Foundation (B).

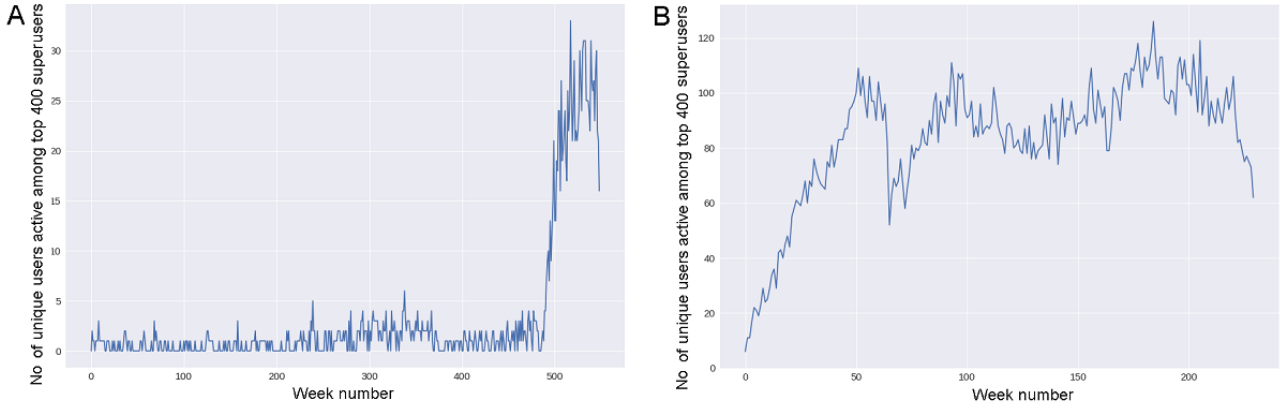


Figure 8 suggests that there was no scarcity of superusers throughout the whole period of observation. In particular, the number of superusers in the Asthma UK community remained stable across almost the entire period until it increased substantially when the forum moved to the HealthUnlocked platform in 2015. Since then about twenty superusers have been active in the forum. On the other hand, in the BLF community the number of unique superusers increased steadily over the first 50 weeks (1 year) since inception (2015), and subsequently there were about 80-100 superusers regularly engaged with the community.

Superusers' Posting Activity

We then investigated whether superusers' posting activity was frequent and regular over time. To this end, for each of the top 5% users by post contribution, calculated cumulatively over the entire observation period, we measured the time interval separating every two subsequent posts to both communities. We then computed the inter-event time distributions for both communities to assess frequency and patterns of activity. Figure 9 suggests that 70% of interposting times were shorter than 3.1 days in the Asthma UK community, while 65% of interposting times in the BLF community were shorter than 1.7 days.

Superusers' Expertise

For each user, a *z*-score was calculated in both communities to gauge the user's expertise (see Data Analysis section). Figure 10 suggests that the more users became active in the communities, the more likely they were to write posts (assumed to be "help-giving" posts) [36,45] than to start new threads (assumed to be "help-seeking" posts). Such a finding might indicate that superusers were also those with the necessary degree of expertise to answer a large number of questions.

Thus, superusers not only play a topologically important role in the communities, but they are also likely to provide the expertise needed to answer queries.

Ego Networks of Superusers

Next, we examine whether the ego networks of different types of users were topologically different, and what generated such differences. Users commonly started a discussion thread by writing a root post (ie, the post at level 1 of the thread). Several users could then directly respond to these posts at level 1, thus creating level-2 posts. More generally, according to the design of the communities, by posting a response to a level–(t) post, users created a level–(t+1) post. There was no limitation to how a post thread could evolve, and therefore to the complexity of the thread hierarchy. Information on post levels was made available through the post metadata. In our analysis, any post at level 2 or higher was classified as a level–2+ post. Here the analysis was restricted to the BLF forum, as the Asthma UK community was significantly smaller with simpler hierarchical levels.

Figure 9. Cumulative distribution function (CDF) of the interposting time for the top 5% of users by post contribution within the Asthma UK (A) and the British Lung Foundation (B) communities.

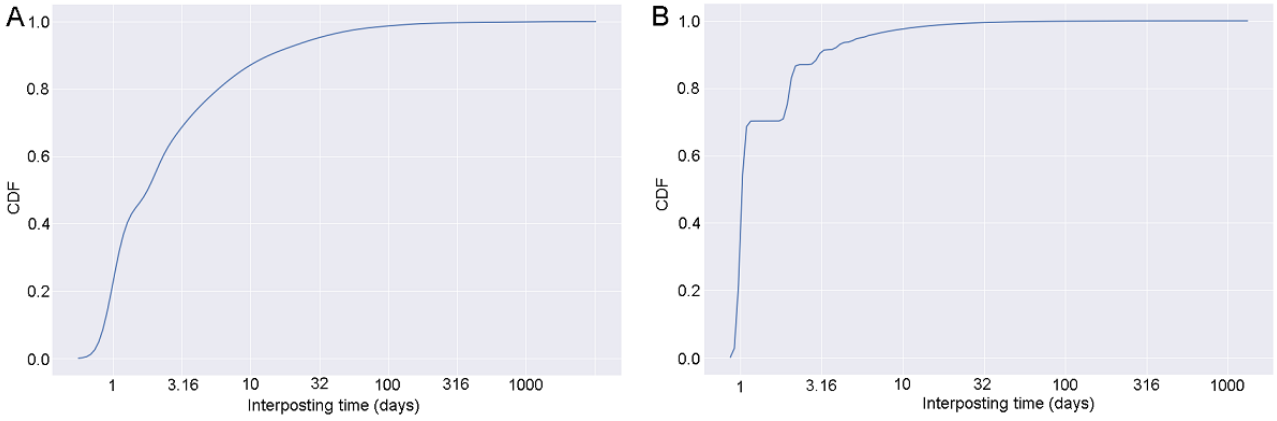


Figure 10. *Z*-score values of all users as a function of the number of posts written in the Asthma UK (A) and British Lung Foundation (B) communities. The top panels represent the normalized distributions of the number of users who wrote various numbers of posts.

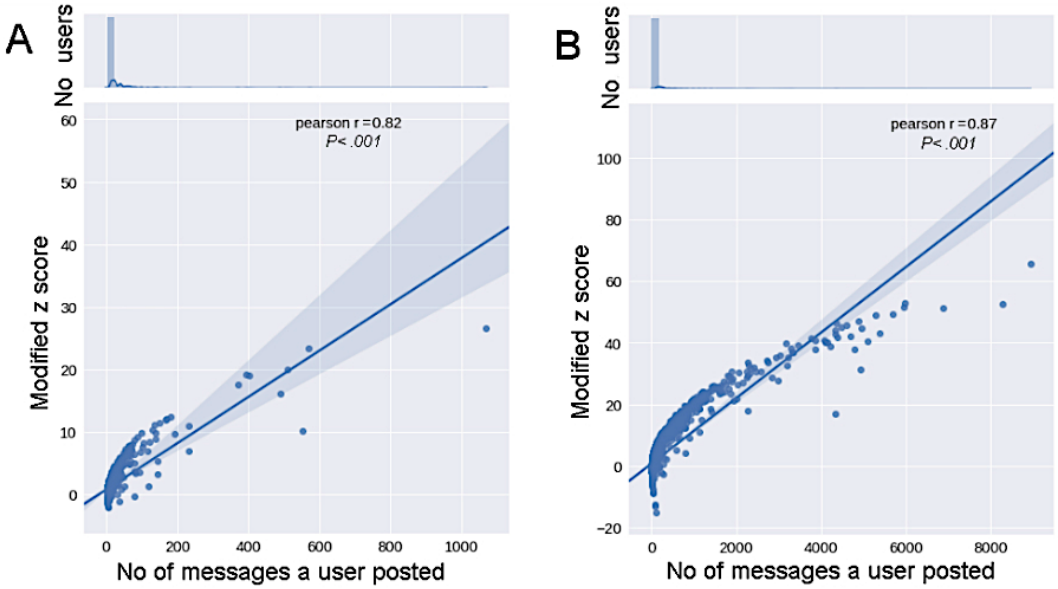
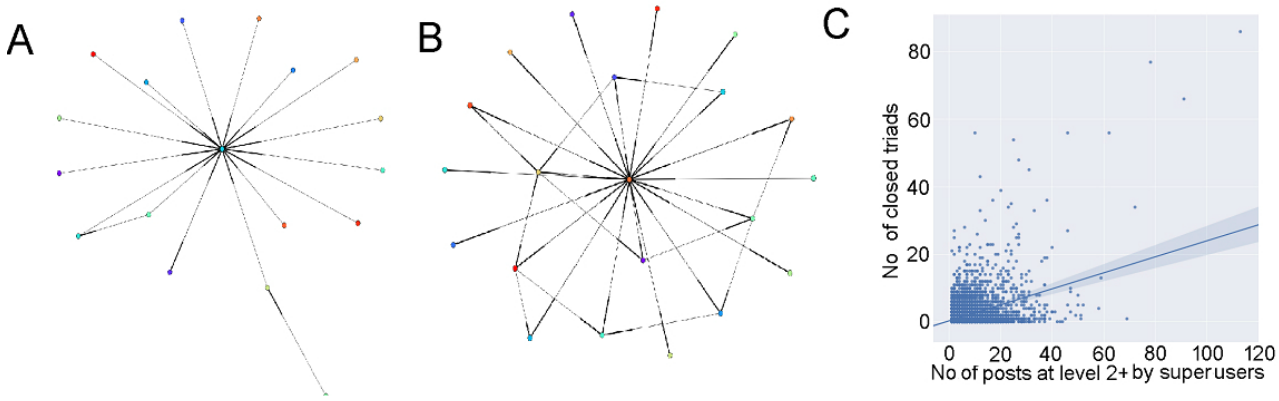


Figure 10A and B show the ego networks of two types of users: one where the help-seeker, called root poster, contributed back multiple times to the thread itself, and the other where this pattern did not happen. In both cases, the thread received similar community engagements in terms of responses from other users. Figure 11B suggests that the highly active root poster developed a more cohesive network, rich in third-party relationships. In this ego network, many alters indeed connected with one another, thus creating closed triads centered on ego. In simple words, these users' posting activity had the effect of making other users talk to each other, thus increasing integration and cohesion within the community. By contrast, the ego network developed by the root poster characterized by a lower contribution to the thread (Figure 11A) had a star-like shape and was rich in structural cleavages between alters. In this ego network, alters were disconnected from each other, and ego acted as the broker enabling indirect connections between alters. In simple words, these users did not favor connections between other users.

By replying to other users' posts, superusers contributed significantly to level 2 or above. Figure 11C shows that there was a significant correlation between the number of triads in an ego network and the number of times ego (the root poster) contributed to the thread itself. The correlation coefficient between the number of triads and the number of posts at level 2 or above written by the top 5% of users by post contribution is 0.44 (P<.001).

When root posters responded back to the posts received, they created a more cohesive network structure. Most of these highly active users were superusers. This suggests that superusers, by posting "help-giving" posts, enabled other users to talk to each other, thus facilitating the formation of ties between them.

Figure 11. Topology of two illustrative ego networks created by a user with low (A) and high (B) posting activity in the British Lung Foundation community. Panel C shows the number of closed triads in ego networks as a function of posting activity of superusers (top 5% of users by post contribution).



Discussion

Summary of Main Findings

In this study, we applied network analysis to two online communities for patients with chronic respiratory conditions to shed light on potential structural mechanisms underlying the role of these communities as scalable, peer-to-peer self-management support intervention systems. We found that the number of users and posts increased steadily over the years in the period of analysis. The majority of users were mutually reachable, either directly or indirectly, and formed a large connected component, which underlies the strength of the network as a means for widespread diffusion of information.

Superusers played a central role in these communities as a result of the characteristics of their posting activity and their constant online engagement. They preferentially replied to posts from peripheral users who were not equally well connected. In doing so, they additionally facilitated tie formation between users. Sensitivity analysis showed that gradual removal of superusers induced the network to collapse. Thus, superusers were responsible for holding the network together and, in particular, for ensuring the emergence of a large connected component. As a result, without superusers, there would be no effective spread of information within the community. Superusers acted as a continuously available resource over time. As users became more active within the community, they became more likely to reply to posts than to ask questions. This suggests that superusers gradually became “experts” providing others with advice and support, which is in agreement with what has recently been suggested by other qualitative studies [6,51].

Strengths and Limitations

Based on social network analysis, this work has started elucidating crucial mechanisms underlying the potential of online health communities to promote effective self-management support interventions, in particular regarding the role of superusers in sustaining and providing integration and cohesion to the network. By analyzing the communities over more than five years, we have shown that superusers are a resource naturally present, able to sustain a network and make it thrive over time. This could prompt future studies to understand their role as a potential scalable health care workforce [1].

Limitations of this study include the lack of demographic and clinical information of participants as well as verification and validation of the information shared online [52], although previous qualitative work by the authors has identified Asthma UK superusers as adolescents with asthma [25]. Moreover, findings were not validated through the semantic analysis of the posts.

We did not investigate the reasons explaining the oscillating number of posts per user per week in the 2 communities, nor the time patterns of posting activity, nor the higher and regular number of posts of BLF users compared with Asthma UK ones. Time patterns of posting activity may reflect the nature of symptoms of the underlying lung conditions (see Figure 4). In particular, the uniformity of posting activity of BLF users might reflect daily self-management activities, whereas the time patterns uncovered for Asthma UK users might reflect self-management activities triggered by episodic exacerbations of symptoms.

More research is also needed to explore the mechanisms sustaining the effectiveness of health online communities and online engagement [53] in terms of the users’ quality of life and, more generally, the generation of beneficial health-related outcomes [54]. The role of superusers in the spread of information within online communities calls for further research to investigate how they can improve quality of information and reduce any potential harm [55]. Future work along these lines will integrate available evidence that incorrect or misleading information is, in many cases, efficiently corrected by peers [6,56]. Moreover, recent research has suggested that leveraging superusers to promote users’ online engagement may not achieve improved health-related outcomes, at least in connection with smoking cessation [57]. More qualitative work should, therefore, shed light on the role of superusers as actual providers of help and advice to other users.

Finally, 90% of people accessing patients’ online communities are passive readers who do not engage in online discussions [44,58]. This means that the number of registered users who post in the forum may represent only 10% of the people who

access the community. However, how this large majority of patients that passively access patients' online communities can benefit from reading others' posts requires further investigation [59]. In particular, it remains unclear whether passive users can improve their self-management and other health-related behaviors, although previous work has shown that participation in online communities can increase passive users' sense of belonging [60]. Change in behaviors of passive readers needs to be fully accounted for to examine the cost-effectiveness of peer-based online support interventions, compared with more traditional intervention tools. Moreover, it remains to be investigated whether there are variations in cost-effectiveness across active users and sub-groups of them with different patterns of social ties [61].

Comparison With Related Work

Previous studies on medical online communities agree that users can benefit from the emotional support as well as the cumulative experiential information provided by others [1,62,63]. The value of online self-management support lies in the availability of co-created experiential knowledge and the presence of distributed health literacy. This enables users to find the information they require to manage their condition, and thus allows them to benefit from the health literacy of others in the network [1].

A qualitative study that was performed on a forum of people with stroke has shown that up to 95% of users' intents for writing posts were met by replies [22]. In agreement with previous reports [45], we found that superusers represented a small proportion of the users in both communities, though they contributed to a considerable proportion of the overall posts. Superusers were members who assumed leadership roles by providing support, advice, and direction to other members [64,65].

This is in qualitative agreement with recent work on an online community for people with stroke, where superusers were shown to play an essential role in nurturing the ability of the forum to provide feedback and identify inappropriate information and health behaviors in the context of secondary prevention medications [6]. Interestingly, a related study using linguistic analysis showed that as users' engagement in the community increased, their use of language changed. For example, it has been documented that the frequency of imperative verbs rose steadily through membership length, as superusers explicitly directed new members to do certain things [51].

Finally, superusers' engagement with the online community and their daily commitment raise questions about what motivates their behavior. Recent work has suggested that their behavior can be motivated by perceived improvements in sense of well-being [4]. Thus, superusers can themselves profit from their engagement with online health communities. However, what remains to be investigated is whether and to what extent spending so much time in online health communities might be detrimental to superusers' self-management.

Implications for Policy, Practice, and Research

As a result of the voluntary basis of users' contributions, self-management support through online health communities offers high potential for cost-effectiveness from the perspective of formal services. Current health care challenges [66] include supporting self-care and management of LTCs. A key to future changes in models of health and social care are the expansion of health services offered locally as well an increasing role for patient self-management of LTCs. Initiatives to improve access to care in the community include expanding health care team to incorporate more allied health care professionals [67]. The benefits of self-management have not been realized through conventional face-to-face channels [18]. Could superusers represent an allied health care workforce, providing a means for health and social care integration? The impact and benefit of this novel approach could be huge and include: (a) increasing the confidence of a large number of people to self-care, (b) reducing demand on general practices [15], emergency care services and hospitals, and (c) saving money within health care systems, and across society as a whole. The potential scale of societal benefits would likely outweigh the opportunity costs associated with the time contributed by users. Understanding the mechanisms underlying effectiveness and uncovering how online communities are organized and evolve are vital preludes to developing and testing effective interventions and are required by the Medical Research Council Complex Interventions Framework [34]. However, little work has addressed this area to date. Although there is evidence that highly engaged users play a role as active help-providers to other users [45], this is to our knowledge the first study showing that superusers in online health communities: (1) are responsible for holding the community together, (2) engage with other users with low posting activity, and (3) indirectly contribute to tie formation between other users.

This work has drawn on social network analysis to uncover fundamental mechanisms underlying the potential of online communities to promote effective self-management support interventions. In particular, our study contributes to a better understanding of the role played by superusers in sustaining and providing integration and cohesion to the network. By analyzing the communities over more than five years, we have shown that superusers can sustain and make the network thrive over time. The presence of both a large connected component and superusers is a crucial feature of successful health communities. It is well known that components are critical for information diffusion [50,68]. Without a large connected component, users would be members of small isolated islands, and information would be unable to flow from one island to another. An online community needs a large component to function effectively. As edges between users are added over time, a large component is likely to emerge [69]. Our work has shown not only that superusers play a critical role in the emergence of a connected component, but also that, even without being "appointed" externally, superusers would emerge as the community grows large enough. Our findings will, therefore, prompt and inform future research interested in understanding superusers' role as a potential scalable health care workforce in online self-management support interventions [1,70].

Moreover, our study has uncovered temporal patterns of posting activity. This will prompt further research aimed at investigating

differences in these patterns across communities using qualitative analysis. This would include the analysis of whether users' intents were met by replies [22] and the potential correlation between the amount of time spent online and improved disease self-management.

Across a variety of empirical domains, it has been documented that hubs (ie, nodes with a disproportionally large number of connections) are valuable resources that help spread information widely and amplify information cascades [71], help design effective vaccination campaigns and selective immunization strategies against disease diffusion and epidemics [72,73], and help improve the system's robustness and vulnerability to random failures [74]. Here we have shown that health online communities are no exception. Our results suggest that superusers indeed represent a crucial resource for the integration and functioning of such communities, which therefore seem to be governed by the same network mechanisms as other real-world networks. This study will, therefore, inform future research interested in uncovering the common organizing principles underpinning a variety of real-world systems.

Conclusions

This study shows that patients' online communities share the same network features as other complex networks across a variety of empirical domains. Our analysis highlighted the special role played by superusers, their topological positions and behavior in the communities. In this sense, our results shed light on the topological mechanisms underlying the ability of patients' online communities to provide self-management support and may, therefore, suggest levers for improving the quality of health care intervention.

At a time when health care services are working beyond capacity and patients are finding it difficult to access care, online communities provide the potential for addressing critical health care challenges. They offer a feasible way for patients with LTCs to find helpful advice and support, and a potentially cost-effective and scalable solution to the vast and rising costs associated with long-term disease management. Even though our results showed that there was no scarcity of superusers throughout the whole period of the study, nonetheless ensuring that such networks will become a core component of illness self-management on a broader scale requires proper research investment leading to randomized control studies and potentially a change in the concept of the health care team.

Acknowledgments

We would like to thank Asthma UK and British Lung Foundation for granting the permission to conduct the study. This study was funded by a Queen Mary University of London Life Science Initiative grant (supported by the Wellcome Trust Institutional Strategic Support Fund). ADS is funded by a National Institute for Health Research Academic Clinical Lectureship.

Authors' Contributions

ADS conceived the study, contributed to the data analysis and interpretation and wrote the manuscript together with PP and SJ. SJ conducted the social network analysis under the guide of PP and NS. PP, NC, SJCT, AP, AS, RD, AA, and MJE are coinvestigators on the study and contributed to the interpretation of findings. CJG contributed to the design of the analysis and interpretation of findings. All authors commented and agreed on the final draft of the submitted manuscript.

Conflicts of Interest

The views expressed are those of the author(s) and not necessarily those of the National Health Service, the National Institute for Health Research or the Department of Health. The funder had no role in study design, data collection, data analysis, data interpretation, the writing of the manuscript, and decision to submit the manuscript for publication. MJE is the cofounder, and chief medical officer of HealthUnlocked and AA is a research officer at HealthUnlocked.

Multimedia Appendix 1

Asthma UK cumulative activity over the analysis time frame.

[MP4 File (MP4 Video), 856KB - jmir_v20i7e238_app1.mp4]

Multimedia Appendix 2

British Lung Foundation cumulative activity over the analysis time frame.

[MP4 File (MP4 Video), 5MB - jmir_v20i7e238_app2.mp4]

References

1. Allen C, Vassilev I, Kennedy A, Rogers A. Long-Term Condition Self-Management Support in Online Communities: A Meta-Synthesis of Qualitative Papers. J Med Internet Res 2016;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]

2. Nesta. Peer Support: What is it and does it work? Summarising evidence from more than 1000 studies, 2015 URL: https://www.nationalvoices.org.uk/sites/default/files/public/publications/peer_support_-_what_is_it_and_does_it_work.pdf [WebCite Cache ID 6zO7rLyaj]
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. Pendry L, Salvatore J. Individual and social benefits of online discussion forums. Computers in Human Behavior 2015;50:211-220.
5. Bartlett YK, Coulson NS. An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication. Patient Educ Couns 2011 Apr;83(1):113-119. [doi: 10.1016/j.pec.2010.05.029] [Medline: 20599338]
6. Izuka NJ, Alexander MAW, Balasooriya-Smeekens C, Mant J, De Simoni A. How do stroke survivors and their carers use practitioners' advice on secondary prevention medications? Qualitative study of an online forum. Fam Pract 2017 Sep 01;34(5):612-620. [doi: 10.1093/fampra/cmx023] [Medline: 28398553]
7. Hobbs WR, Burke M, Christakis NA, Fowler JH. Online social integration is associated with reduced mortality risk. Proc Natl Acad Sci U S A 2016 Dec 15;113(46):12980-12984 [FREE Full text] [doi: 10.1073/pnas.1605554113] [Medline: 27799553]
8. Panagioti M, Richardson G, Small N, Murray E, Rogers A, Kennedy A, et al. Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. BMC Health Serv Res 2014;14:356 [FREE Full text] [doi: 10.1186/1472-6963-14-356] [Medline: 25164529]
9. Taylor SJ, Pinnock H, Epiphaniou E, Pearce G, Parke HL, Schwappach A, et al. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions: PRISMS – Practical systematic Review of Self-Management Support for long-term conditions. Health Serv Deliv Res 2014;2(53):1-580. [Medline: 25642548]
10. Kirk S, Milnes L. An exploration of how young people and parents use online support in the context of living with cystic fibrosis. Health Expect 2015 Feb 17:309-321. [doi: 10.1111/hex.12352] [Medline: 25691209]
11. Stewart M, Letourneau N, Masuda JR, Anderson S, McGhan S. Online solutions to support needs and preferences of parents of children with asthma and allergies. J Fam Nurs 2011;17(3):357-379. [doi: 10.1177/1074840711415416] [Medline: 21813815]
12. Matura LA, McDonough A, Aglietti LM, Herzog JL, Gallant KA. A virtual community: concerns of patients with pulmonary hypertension. Clin Nurs Res 2013 May;22(2):155-171. [doi: 10.1177/1054773812462867] [Medline: 23093554]
13. Wentzer HS, Bygholm A. Narratives of empowerment and compliance: studies of communication in online patient support groups. Int J Med Inform 2013 Dec;82(12):e386-e394. [doi: 10.1016/j.ijmedinf.2013.01.008] [Medline: 23523082]
14. London: Department of Health. 2012. Long Term Conditions Compendium of Information, 3rd ed URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216528/dh_134486.pdf [accessed 2018-06-29] [WebCite Cache ID 70XxR0HFn]
15. De Simoni A, Griffiths CJ, Taylor SJ. Improving access to primary care: can online communities contribute? Br J Gen Pract 2016 Nov;66(652):559 [FREE Full text] [doi: 10.3399/bjgp16X687721] [Medline: 27789494]
16. The expert patient: a new approach to chronic disease management in the 21st century. Londontationery Office, 2001 URL: http://webarchive.nationalarchives.gov.uk/20120511062115/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4018578.pdf [accessed 2018-05-13] [WebCite Cache ID 6zODDgSnD]
17. Griffiths C, Foster G, Ramsay J, Eldridge S, Taylor S. How effective are expert patient (lay led) education programmes for chronic disease? BMJ 2007 Jun 16;334(7606):1254-1256 [FREE Full text] [doi: 10.1136/bmj.39227.698785.47] [Medline: 17569933]
18. Greenhalgh T. Patient and public involvement in chronic illness: beyond the expert patient. BMJ 2009 Feb 17;338:b49. [Medline: 19223339]
19. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004 Jun;7(3):321-326. [doi: 10.1089/1094931041291295] [Medline: 15257832]
20. Fox S. 2011 URL: http://www.pewInternet.org/Reports/2011/Social-Life-of-Health-Info/Summary-of-Findings.aspx [accessed 2018-05-28] [WebCite Cache ID 6zO95LhSc]
21. Armstrong A, Hagel J. Knowledge and communities. 2000. The real value of online communities URL: https://linkinghub.elsevier.com/retrieve/pii/B9780750672931500093 [accessed 2018-06-29] [WebCite Cache ID 70XxvOQ91]
22. De Simoni A, Shanks A, Balasooriya-Smeekens C, Mant J. Stroke survivors and their families receive information and support on an individual basis from an online forum: descriptive analysis of a population of 2348 patients and qualitative study of a sample of participants. BMJ Open 2016 Apr 06;6(4):e010501 [FREE Full text] [doi: 10.1136/bmjopen-2015-010501] [Medline: 27053271]
23. Coulson NS. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. Cyberpsychol Behav 2005 Dec;8(6):580-584. [doi: 10.1089/cpb.2005.8.580] [Medline: 16332169]
24. Shoebotham A, Coulson NS. Therapeutic Affordances of Online Support Group Use in Women With Endometriosis. J Med Internet Res 2016 May 09;18(5):e109 [FREE Full text] [doi: 10.2196/jmir.5548] [Medline: 27160641]

25. De Simoni A, Horne R, Fleming L, Bush A, Griffiths C. What do adolescents with asthma really think about adherence to inhalers? Insights from a qualitative analysis of a UK online forum. BMJ Open 2017 Jun 13;7(6):e015245 [FREE Full text] [doi: 10.1136/bmjopen-2016-015245] [Medline: 28615272]
26. Ennis-O'Connor M. BMJ Blog. 2014. How online patient communities are changing the face of cancer care URL: http://blogs.bmj.com/ebn/2014/03/03/how-online-patient-communities-are-changing-the-face-of-cancer-care/ [accessed 2018-01-15] [WebCite Cache ID 6zO9uFksP]
27. Ali K, Farrer L, Gulliver A, Griffiths KM. Online Peer-to-Peer Support for Young People With Mental Health Problems: A Systematic Review. JMIR Ment Health 2015;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
28. Carron-Arthur B, Cunningham J, Griffiths K. Describing the distribution of engagement in an Internet support group by post frequency: A comparison of the 90-9-1 Principle and Zipf's Law. Internet Interventions 2014;1(4):165-168 [FREE Full text]
29. van Mierlo T, Hyatt D, Ching AT. Mapping Power Law Distributions in Digital Health Social Networks: Methods, Interpretations, and Practical Implications. J Med Internet Res 2015;17(6):e160 [FREE Full text] [doi: 10.2196/jmir.4297] [Medline: 26111790]
30. van Mierlo T, Hyatt D, Ching AT. Employing the Gini coefficient to measure participation inequality in treatment-focused Digital Health Social Networks. Netw Model Anal Health Inform Bioinform 2016;5(1):32 [FREE Full text] [doi: 10.1007/s13721-016-0140-7] [Medline: 27840788]
31. Carron-Arthur B, Reynolds J, Bennett K, Bennett A, Cunningham JA, Griffiths KM. Community Structure of a Mental Health Internet Support Group: Modularity in User Thread Participation. JMIR Ment Health 2016;3(2):e20 [FREE Full text] [doi: 10.2196/mental.4961] [Medline: 27242012]
32. Gopalsamy R, Semenov A, Pasiliao E, McIntosh S, Nikolaev A. Engagement as a Driver of Growth of Online Health Forums: Observational Study. J Med Internet Res 2017 Aug 29;19(8):e304 [FREE Full text] [doi: 10.2196/jmir.7249] [Medline: 28851677]
33. Urbanoski K, van Mierlo T, Cunningham J. Investigating Patterns of Participation in an Online Support Group for Problem Drinking: a Social Network Analysis. Int.J. Behav. Med 2016 Aug 22;24(5):703-712. [doi: 10.1007/s12529-016-9591-6]
34. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [Medline: 18824488]
35. Burt RS. The network structure of social capital. Research in Organizational Behavior 2000;22:345-423.
36. Zhang J, Ackerman M, Lada A. Expertise networks in online communities: structure and algorithms. In: Proceedings of the 16th international conference on World Wide Web. 2007 Presented at: WWW 2007; May 8-12 2007; Banff, Alberta, Canada. [doi: 10.1145/1242572.1242603]
37. Zhao K, Wang X, Cha S, Cohn AM, Papandonatos GD, Amato MS, et al. A Multirelational Social Network Analysis of an Online Health Community for Smoking Cessation. J Med Internet Res 2016 Aug 25;18(8):e233 [FREE Full text] [doi: 10.2196/jmir.5985] [Medline: 27562640]
38. Song. Collaborative friendship networks in online healthcare communities: an exponential random graph model analysis. In: Smart Health: International Conference, ICSH 2014, July 10-11, 2014. Proceedings (Lecture Notes in Computer Science). Beijing, China: Springer; 2014.
39. HealthUnlocked. The social network for health. URL: https://healthunlocked.com/ [accessed 2018-06-29] [WebCite Cache ID 70Y10rppl]
40. Wasserman S, Faust K, Granovetter M. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press; 1994.
41. Hawe P, Webster C, Shiell A. A glossary of terms for navigating the field of social network analysis. J Epidemiol Community Health 2004 Dec;58(12):971-975 [FREE Full text] [doi: 10.1136/jech.2003.014530] [Medline: 15547054]
42. Panzarasa P, Bonaventura M. Emergence of long-range correlations and bursty activity patterns in online communication. Phys Rev E Stat Nonlin Soft Matter Phys 2015 Dec;92(6):062821. [doi: 10.1103/PhysRevE.92.062821] [Medline: 26764758]
43. Colizza V, Flammini A, Serrano MA, Vespignani A. Detecting rich-club ordering in complex networks. Nat Phys 2006 Jan 15;2(2):110-115. [doi: 10.1038/nphys209]
44. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
45. 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]
46. Panzarasa P, Bonaventura M. Emergence of long-range correlations and bursty activity patterns in online communication. Phys Rev E Stat Nonlin Soft Matter Phys 2015 Dec;92(6):062821. [doi: 10.1103/PhysRevE.92.062821] [Medline: 26764758]
47. Opsahl T, Colizza V, Panzarasa P, Ramasco JJ. Prominence and control: the weighted rich-club effect. Phys Rev Lett 2008 Oct 17;101(16):168702. [doi: 10.1103/PhysRevLett.101.168702] [Medline: 18999722]
48. Zhou S, Mondragon R. The Rich-Club Phenomenon in the Internet Topology. IEEE Commun. Lett 2004 Mar;8(3):180-182. [doi: 10.1109/lcomm.2004.823426]
49. Newman M. Newman. In: Networks: An Introduction. Oxford: Oxford University Press, USA; 2010.

50. Yagan O, Qian D, Zhang J, Cochran D. Conjoining Speeds up Information Diffusion in Overlaying Social-Physical Networks. IEEE J. Select. Areas Commun 2013 Jun;31(6):1038-1048. [doi: 10.1109/JSAC.2013.130606]
51. McDonald D, Woodward-Kron R. Member roles and identities in online support groups: Perspectives from corpus and systemic functional linguistics. Discourse & Communication 2016;10(2):157-175 [FREE Full text]
52. van Mierlo T, Li X, Hyatt D, Ching AT. Demographic and Indication-Specific Characteristics Have Limited Association With Social Network Engagement: Evidence From 24,954 Members of Four Health Care Support Groups. J Med Internet Res 2017 Feb 17;19(2):e40 [FREE Full text] [doi: 10.2196/jmir.6330] [Medline: 28213340]
53. Iriberri A, Leroy G. A life-cycle perspective on online community success. ACM Comput. Surv 2009 Feb 01;41(2):1-29. [doi: 10.1145/1459352.1459356]
54. Park A, Conway M. Longitudinal Changes in Psychological States in Online Health Community Members: Understanding the Long-Term Effects of Participating in an Online Depression Community. J Med Internet Res 2017 Mar 20;19(3):e71 [FREE Full text] [doi: 10.2196/jmir.6826] [Medline: 28320692]
55. Mo PKH, Coulson NS. Are online support groups always beneficial? A qualitative exploration of the empowering and disempowering processes of participation within HIV/AIDS-related online support groups. Int J Nurs Stud 2014 Jul;51(7):983-993. [doi: 10.1016/j.ijnurstu.2013.11.006] [Medline: 24342449]
56. Esquivel A, Meric-Bernstam F, Bernstam EV. Accuracy and self correction of information received from an internet breast cancer list: content analysis. BMJ 2006 Apr 22;332(7547):939-942 [FREE Full text] [doi: 10.1136/bmj.38753.524201.7C] [Medline: 16513686]
57. Graham AL, Papandonatos GD, Cha S, Erar B, Amato MS, Cobb NK, et al. Improving Adherence to Smoking Cessation Treatment: Intervention Effects in a Web-Based Randomized Trial. Nicotine Tob Res 2017 Mar 01;19(3):324-332. [doi: 10.1093/ntr/ntw282] [Medline: 27836982]
58. Burri M, Baujard V, Etter J. A qualitative analysis of an internet discussion forum for recent ex-smokers. Nicotine Tob Res 2006 Dec;8 Suppl 1:S13-S19. [Medline: 17491166]
59. Mo PKH, Coulson NS. Empowering processes in online support groups among people living with HIV/AIDS: A comparative analysis of lurkers and posters. Computers in Human Behavior 2010;26(5):1183-1193 [FREE Full text]
60. Nonnecke B, Preece J. Shedding light on lurkers in online communities. In: Ethnographic Studies in Real and Virtual Environments: Inhabited Information Spaces and Connected Communities. Edinburgh: Ed. K. Buckner; 1999:123-128.
61. Konchak C, Prasad K. Incorporating social network effects into cost-effectiveness analysis: a methodological contribution with application to obesity prevention. BMJ Open 2012 Oct;2(6):1-11 [FREE Full text] [doi: 10.1136/bmjopen-2012-001038] [Medline: 23117559]
62. Loane SS, D'Alessandro S. Peer-to-peer value through social capital in an online motor neuron disease community. J Nonprofit Pub Sect Mark 2013;25(2):164-185 [FREE Full text]
63. Matura LA, McDonough A, Aglietti LM, Herzog JL, Gallant KA. A virtual community: concerns of patients with pulmonary hypertension. Clin Nurs Res 2013 May;22(2):155-171. [doi: 10.1177/1054773812462867] [Medline: 23093554]
64. van Mierlo T, Voci S, Lee S, Fournier R, Selby P. Superusers in social networks for smoking cessation: analysis of demographic characteristics and posting behavior from the Canadian Cancer Society's smokers' helpline online and StopSmokingCenter.net. J Med Internet Res 2012;14(3):e66 [FREE Full text] [doi: 10.2196/jmir.1854] [Medline: 22732103]
65. Carron-Arthur B, Ali K, Cunningham JA, Griffiths KM. From Help-Seekers to Influential Users: A Systematic Review of Participation Styles in Online Health Communities. J Med Internet Res 2015;17(12):e271 [FREE Full text] [doi: 10.2196/jmir.4705] [Medline: 26627369]
66. NHS England. 2017. NHS Five year forward view URL: https://www.england.nhs.uk/wp-content/uploads/2017/03/NEXT-STEPS-ON-THE-NHS-FIVE-YEAR-FORWARD-VIEW.pdf [accessed 2018-05-13] [WebCite Cache ID 6zOC6wCYw]
67. Primary Care Workforce Commission. Health Education England. 2015. The future of primary care: creating teams for tomorrow URL: https://www.hee.nhs.uk/sites/default/files/documents/The%20Future%20of%20Primary%20Care%20report.pdf [WebCite Cache ID 6zOCBmweh]
68. Lambiotte R, Panzarasa P. Communities, knowledge creation, and information diffusion. Journal of Informetrics 2009 Jul;3(3):180-190. [doi: 10.1016/j.joi.2009.03.007]
69. Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. 2005 Presented at: KDD '05 Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining; August 21–24, 2005; Chicago p. 177-187 URL: https://www.cs.cornell.edu/home/kleinber/kdd05-time.pdf
70. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting Behavioral Interventions for Social Media Delivery. J Med Internet Res 2016;18(1):e24 [FREE Full text] [doi: 10.2196/jmir.5086] [Medline: 26825969]
71. Bakshy E, Hofman J, Mason A, Watts D. Everyone's an influencer: quantifying influence on Twitter. 2011 Presented at: Proceedings of the fourth ACM International Conference on Web searchdata mining (WSDM 11). New York: ACM; February 09 - 12, 2011; Hong Kong, China p. 65-74. [doi: 10.1145/1935826.1935845]
72. Dezso Z, Barabási A. Halting viruses in scale-free networks. Phys Rev E Stat Nonlin Soft Matter Phys 2002 May;65(5 Pt 2):055103. [doi: 10.1103/PhysRevE.65.055103] [Medline: 12059627]

73. Pastor-Satorras R, Vespignani A. Immunization of complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 2002 Mar;65(3 Pt 2A):036104. [doi: 10.1103/PhysRevE.65.036104] [Medline: 11909162]
74. Paul G, Sreenivasan S, Stanley HE. Resilience of complex networks to random breakdown. Phys Rev E Stat Nonlin Soft Matter Phys 2005 Nov;72(5 Pt 2):056130. [doi: 10.1103/PhysRevE.72.056130] [Medline: 16383711]

Abbreviations

BLF: British Lung Foundation
COPD: chronic obstructive pulmonary disease
LTC: long-term condition

Edited by G Eysenbach; submitted 26.01.18; peer-reviewed by T van Mierlo, G Papandonatos, K Tingay; comments to author 15.02.18; revised version received 10.04.18; accepted 12.05.18; published 11.07.18.

Please cite as:
Joglekar S, Sastry N, Coulson NS, Taylor SJC, Patel A, Duschinsky R, Anand A, Jameson Evans M, Griffiths CJ, Sheikh A, Panzarasa P, De Simoni A
How Online Communities of People With Long-Term Conditions Function and Evolve: Network Analysis of the Structure and Dynamics of the Asthma UK and British Lung Foundation Online Communities
J Med Internet Res 2018;20(7):e238
URL: http://www.jmir.org/2018/7/e238/
doi:10.2196/jmir.9952
PMID:29997105

Original Paper

The Association Between Increased Levels of Patient Engagement With an Internet Support Group and Improved Mental Health Outcomes at 6-Month Follow-Up: Post-Hoc Analyses From a Randomized Controlled Trial

Emily M Geramita[1], MD, PhD; Bea Herbeck Belnap[2,3], Dr Biol Hum; Kaleab Z Abebe[1], PhD; Scott D Rothenberger[1], PhD; Armando J Rotondi[3,4], PhD; Bruce L Rollman[3], MPH, MD

[1]Division of General Internal Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA, United States

[2]Department of Psychosomatic Medicine and Psychotherapy, University of Göttingen Medical Center, Göttingen, Germany

[3]Center for Behavioral Health and Smart Technology, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

[4]Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, PA, United States

Corresponding Author:
Bruce L Rollman, MPH, MD
Center for Behavioral Health and Smart Technology
University of Pittsburgh School of Medicine
230 McKee Place, Suite 600
Pittsburgh, PA, 15213
United States
Phone: 1 412 692 2659
Fax: 1 412 692 4838
Email: rollmanbl@upmc.edu

Abstract

Background: We recently reported that depressed and anxious primary care patients randomized to a moderated internet support group (ISG) plus computerized cognitive behavioral therapy (cCBT) did not experience improvements in depression and anxiety over cCBT alone at 6-month follow-up.

Objective: The 1% rule posits that 1% of participants in online communities generate approximately 90% of new user-created content. The aims of this study were to apply the 1% rule to categorize patient engagement with the ISG and identify whether any patient subgroups benefitted from ISG use.

Methods: We categorized the 302 patients randomized to the ISG as: superusers (3/302, 1.0%), top contributors (30/302, 9.9%), contributors (108/302, 35.8%), observers (87/302, 28.8%) and those who never logged in (74/302, 24.5%). We then applied linear mixed models to examine associations between engagement and 6-month changes in health-related quality of life (HRQoL; Short Form Health Survey Mental Health Component, SF-12 MCS) and depression and anxiety symptoms (Patient-Reported Outcomes Measurement Information System, PROMIS).

Results: At baseline, participant mean age was 42.6 years, 81.1% (245/302) were female, and mean Patient Health Questionnaire (PHQ-9), Generalized Anxiety Disorder scale (GAD-7), and SF-12 MCS scores were 13.4, 12.6, and 31.7, respectively. Of the 75.5% (228/302) who logged in, 61.8 % (141/228) created ≥1 post (median 1, interquartile range, IQR 0-5); superusers created 42.3 % (630/1488) of posts (median 246, IQR 78-306), top contributors created 34.6% (515/1488; median 11, IQR 10-18), and contributors created 23.1 % (343/1488; median 3, IQR 1-5). Compared to participants who never logged in, the combined superuser + top contributor subgroup (n=33) reported 6-month improvements in anxiety (PROMIS: –11.6 vs –7.8; *P*=.04) and HRQoL (SF-12 MCS: 16.1 vs 10.1; *P*=.01) but not in depression. No other subgroup reported significant symptom improvements.

Conclusions: Patient engagement with the ISG was more broadly distributed than predicted by the 1% rule. The 11% of participants with the highest engagement levels reported significant improvements in anxiety and HRQoL.

Trial Registration: ClinicalTrials.gov NCT01482806; https://clinicaltrials.gov/ct2/show/NCT01482806 (Archived by WebCite at http://www.webcitation.org/708Bjlge9).

(J Med Internet Res 2018;20(7):e10402) doi:10.2196/10402

KEYWORDS

internet support group; patient engagement; anxiety; depression

Introduction

Background

Internet support groups (ISGs) are specialized social media websites that connect individuals with common health conditions and provide a forum for peers to exchange information, resources, and support [1,2]. While ISGs for mental health conditions have become increasingly common [3], randomized trials [4,5] and systematic reviews [6-8] find they have mixed benefits for reducing psychologic distress. In a recent randomized controlled trial, we reported that providing depressed and anxious primary care patients with access to a moderated ISG in addition to a computerized cognitive behavioral therapy (cCBT) program provided no additional intent-to-treat benefit in patients' health-related quality of life (HRQoL) or mood and anxiety symptoms over the cCBT program alone at 6-month follow-up, although cCBT was more effective than primary care physicians' (PCPs) usual care [9] (NCT01482806). These null findings raise questions about whether any subgroups of ISG members may have benefitted differentially from the ISG based on their level of engagement.

One approach to classify engagement with an online community is the 1% rule [10,11]. Adapted from the digital marketing literature, the 1% rule posits that 1% of online community members (superusers) create approximately 90% of user-generated content, approximately 10% of members (contributors) create less than 10% of the remaining content, and 90% of members (observers) rarely contribute but mainly observe activity. A recent observational study replicated the 1% rule in 4 large ISGs for individuals with addiction and mood disorders [12] and found that participants' demographic and disease-specific characteristics were not associated with their level of engagement with these online communities [13].

Goal of This Study

Very little work has been done to investigate the relationship between level of ISG engagement and clinical outcomes for treating depression and anxiety or any other mental health condition in primary care [5,14]. Therefore, to classify the patients randomly assigned to our trial's ISG arm by their level of engagement, we applied the 1% rule based on the number of posts they created on the ISG. We then conducted post hoc analyses to compare these engagement level subgroups with patients randomly assigned to the ISG who never logged in to examine whether any patient subgroup benefitted from participating in our online community.

Methods

Study Setting, Patient Eligibility, and Randomization and Experimental Conditions

The protocol for the Online Treatment for Mood and Anxiety Disorders Trial was approved by the University of Pittsburgh's Institutional Review Board and detailed in the trial's primary outcomes report [9]. Briefly, PCPs in 26 southwestern Pennsylvania practices referred patients with a Generalized Anxiety Disorder scale (GAD-7) [15] or Patient Health Questionnaire (PHQ-9) [16] score ≥10, indicating moderately severe anxiety or depression symptoms, between August 2012 and September 2014. We randomized 704 protocol-eligible participants to either (1) care manager–guided access to the 8-session "Beating the Blues" cCBT program designed to provide users with basic CBT skills [17] (cCBT-only; 301/704), (2) cCBT plus additional access to our password-protected and moderated ISG (ISG+cCBT; 302/704), or (3) their PCP's usual care (101/704). All study arms had similar baseline sociodemographic and clinical characteristics [9]. Analyses in this report focus solely on the 302 participants assigned to the ISG+cCBT arm.

Care Manager Support

Following randomization, the care manager exclusively assigned to the ISG+cCBT arm contacted each participant via telephone to provide basic psychoeducation and encourage them to start the cCBT program and log in to the ISG. Later, he contacted participants (we define these as care manager contacts) via email, text, and telephone to promote adherence with the cCBT program and treatment recommendations, including suggestions to access various resources on the ISG. The care manager presented each participant's progress to the study PCP, psychiatrist, and psychologist at a weekly case review meeting [18].

Internet Support Group

We used WordPress (Automattic Inc) software to create our password-protected ISG that was accessible via computer or smartphone (Figure 1). The ISG featured moderated discussion boards created by the care manager–ISG moderator and study participants. The study team used an iterative process to decide on the initial discussion board topics, with a focus on common challenges faced by patients with depression and anxiety (eg, managing symptoms, discussing mental health issues with friends, common triggers). The ISG also curated links to external resources including local $4 generic pharmacy programs; find-a-therapist; crisis hotlines; brief YouTube videos on insomnia, nutrition, exercise, and other topics; our electronic medical record system's patient portal; and the cCBT program (Multimedia Appendix 1).

To preserve confidentiality, we assigned members usernames and regularly reminded them not to post any self-identifying information or photographs. Additionally, a study investigator logged in to the ISG daily to review new posts for suicidal thoughts and other potentially inappropriate content. Participants were also able to flag comments for review by the ISG moderator and possible removal.

Figure 1. Screenshot of our internet support group homepage (ottrial.pitt.edu).

Engagement with the Internet Support Group

We provided participants with password-protected access to the ISG approximately 3 months after the start of subject enrollment once the first 25 patients were randomized to the ISG arm to promote user-generated activity. Afterward, we provided participants with ISG access shortly after randomization.

Participants created content on the ISG discussion boards by either initiating a new discussion thread or commenting on an ongoing thread (posts). On most weeks, the care manager–ISG moderator also initiated new discussion threads on such topics as coping with mental health symptoms, talking about depression and anxiety with friends, stressors (eg, holidays, work-life balance), and lifestyle challenges (eg, healthy diet, losing weight, exercise).

Although we encouraged participants to log in and post on the ISG throughout their 6-month intervention phase, we did not require them to do so. Still, we took several measures to encourage participants to log in and post by featuring status indicators on their profiles and posts (eg, stars and likes), emailing notifications of new ISG activities and posts, highlighting new posts on their homepage based on their past ISG activity, inviting participants to serve as guest moderators, and holding various contests that promoted logging in and posting.

Assessments

Following confirmation of protocol eligibility and consent, a study assessor collected sociodemographic and clinical information from our study practices' electronic medical record system and from the participant, ascertained self-identified race, and administered the Primary Care Evaluation of Mental Disorders (PRIME-MD) Anxiety and Mood Modules to establish a psychiatric diagnosis [19], the 12-Item Short Form Health Survey Mental Components Score (SF-12 MCS) to measure HRQoL [20], and the Patient-Reported Outcomes Measurement Information System (PROMIS) Depression and Anxiety short forms to measure depression and anxiety symptom levels [21]. Later, an assessor who was blinded to participants' randomization assignment telephoned participants to readminister the PROMIS and SF-12 MCS at 3 months and at the 6-month primary outcome time point.

We obtained counts of unique patient log-ins and posts from the logs of the server that hosted the ISG. We defined a post as an entry that initiated a new discussion thread or added an entry to an existing discussion thread, and we summarized the number of posts each participant made to arrive at a total.

Classification of Internet Support Group Engagement

Using the 1% rule as our starting point, we classified participants into subgroups by level of engagement as measured by the total

number of posts each created during the first 6 months after randomization (top 1% of posters, next 9%, and remaining 90%) [10,11]. Given our interest in identifying the gradient of participant engagement, we further classified participants into the following subgroups: superusers (top 1%), top contributors (next 9%), contributors (made at least 1 post), observers (logged in at least once but never posted), and those who never logged in. Since several participants between the 9th and 11th percentiles made the same number of posts, we reclassified our top contributors as the next highest 10% of posters after superusers, rather than the next 9%.

Statistical Analysis

We calculated the baseline sociodemographic and clinical characteristics across the 5 ISG engagement groups using percentages, means and standard deviations, and medians and interquartile ranges (IQR), and we made group comparisons using analysis of variance and chi-square tests. As we had only 3 superusers, we grouped them with the 30 top contributors for all analyses to conduct more meaningful comparisons.

We used linear mixed models for each of the clinical outcomes (SF-12 MCS, PROMIS Depression, PROMIS Anxiety) that included fixed effects for engagement subgroup, time, group-by-time interaction, education, self-identified race, gender, and random effects for participants. We also compared the 6-month change in HRQoL and depression and anxiety symptoms between participants who were assigned to the ISG arm (ISG+cCBT) but never logged in and participants in our combined superuser + top contributor subgroup. All analyses were conducted using SAS version 9.4 (SAS Institute Inc).

Results

Baseline Sociodemographic and Clinical Characteristics

At baseline (Table 1), the 302 participants randomized to the ISG+cCBT arm reported moderately severe depression (PHQ-9 mean 13.4, SD 4.7) and anxiety symptoms (GAD-7 mean 12.6, SD 4.5) and low HRQoL (SF-12 MCS mean 31.7, SD 9.4). They had a mean age of 42.6 years, 81.1% (245/302) were female, and 47.7% (144/302) had at least a college education.

While each engagement subgroup was predominately female, white, and had comorbid depression and anxiety, reflecting the overall composition of our study cohort, participants who were female, white, and college educated were more likely to be in the superusers + top contributors subgroup (eg, ≥4-year college education: 70%, 23/33 superusers + top contributors vs 36%, 27/74 of the never log-ins; Table 1).

Distribution of Engagement

Seventy-five percent of participants (228/302) logged in to the ISG at least once during their 6-month intervention phase, for a total of 2041 log-ins. Of those, the median number of log-ins per participant was 4 (IQR 2-9.5; range 1-214). Participants created 1488 posts over the 6-month intervention phase, and 61.8% (141/228) made at least 1 post (median posts per participant: 1, IQR 0-5).

As expected, the mean number of log-ins and posts differed widely across engagement subgroups (P<.001, Tables 2 and 3). However, the distribution of posts in our sample was less skewed than predicted by the 1% rule, with superusers making 42.3% (630/1488) of posts (median 246, IQR 78-306), top contributors 34.6% (515/1488, median 11, IQR 10-18) and contributors 23.1% (343/1488, median 3, IQR 1-5). Moreover, only 28.8% (87/302) of participants in the ISG were classified as observers (ie, they logged in to the site at least once but never posted).

Process Measures of Care

Overall, the mean number of cCBT sessions completed was 5.5 (SD 2.7), and 35.8% (108/302) completed all 8 cCBT sessions. Across engagement subgroups, participants who created more posts also completed more cCBT sessions (P<.001) and had more care manager contacts (P<.001; Tables 2 and 3).

Mental Health Outcomes at 6 Months

After adjusting for gender, race, and education level, all engagement subgroups reported similar improvements in symptoms at 6-month follow-up regardless of level of engagement with the ISG (Tables 4 and 5). Furthermore, compared to participants who never logged in to the ISG, the combined superusers + top contributors subgroup reported a greater improvement in HRQoL (mean Δ SF-12 MCS: 16.1, SE 1.9 vs 10.1, SE 1.3, P=.01) and anxiety symptoms (mean Δ PROMIS T-score: –11.6, SE 1.5 vs –7.8, SE 1.0, P=.04); we did not observe a similar improvement in depression symptoms.

Table 1. Baseline sociodemographic and psychiatric characteristics by engagement level.

Characteristic	Overall (n=302)	Superusers + top contributors (n=33)	Contributors (n=108)	Observers (n=87)	Never log-ins (n=74)	*P* value[a]
Age, mean (SD)	42.6 (14.4)	40.9 (13.3)	41.9 (14.4)	43.0 (14.0)	43.9 (15.5)	.72
Female, n (%)	245 (81.1)	31 (94)	88 (81.4)	63 (72)	63 (85)	.04[b]
White race, n (%)	242 (80.1)	29 (88)	94 (87.0)	65 (75)	54 (73)	.04[b]
≥4-year college degree, n (%)	144 (47.7)	23 (70)	54 (50.0)	40 (46)	27 (36)	.02
Married or living with partner, n (%)	120 (39.7)	18 (55)	42 (38.9)	36 (41)	24 (32)	.38
Employed, n (%)	204 (67.6)	22 (67)	75 (69.4)	62 (71)	45 (61)	.52
Psychiatric[c] diagnosis, n (%)						.75[b]
Major depression only	63 (21.6)	8 (24)	21 (19.4)	21 (24)	13 (18)	
Generalized anxiety disorder only	22 (7.5)	3 (9)	5 (4.6)	7 (8)	7 (9)	
Both depression and anxiety	207 (70.9)	22 (67)	78 (72.2)	57 (66)	50 (68)	
PHQ-9[d], mean (SD)[e]	13.4 (4.7)	12.3 (5.5)	14.0 (4.4)	13.2 (4.6)	13.3 (4.7)	.27
GAD-7[f], mean (SD)[e]	12.6 (4.5)	13.3 (4.8)	12.8 (4.6)	12.6 (4.7)	12.1 (3.8)	.62
PROMIS[g] Depression T-score, mean (SD)	62.0 (6.3)	61.8 (6.7)	62.4 (6.1)	62.0 (6.5)	61.5 (6.1)	.79
PROMIS Anxiety T-score, mean (SD)	65.8 (6.2)	66.9 (6.7)	66.0 (6.5)	65.6 (5.4)	65.2 (6.2)	.57
SF-12 MCS[h], mean (SD)	31.7 (9.4)	31.4 (8.9)	30.9 (9.2)	31.2 (8.5)	33.6 (10.9)	.25
Depression/anxiety medication use in past year, n (%)	236 (78.1)	26 (79)	87 (80.6)	68 (78)	55 (74)	.66[b]
Mental health therapist visit in past year, n (%)	59 (19.5)	10 (30)	18 (16.7)	19 (22)	12 (16)	.35

[a]*P* value represents comparison of the 4 engagement level groups.

[b]*P* value from Fisher exact test.

[c]10 participants did not meet diagnostic criteria for depression or anxiety on the Primary Care Evaluation of Mental Disorders; these participants were not included in the denominator when calculating the percentage with each diagnosis.

[d]PHQ-9: Patient Health Questionnaire.

[e]n=30 in Superusers and Top contributors group.

[f]GAD-7: Generalized Anxiety Disorder scale.

[g]PROMIS: Patient-Reported Outcomes Measurement Information System.

[h]SF-12 MCS: Short Form Health Survey Mental Components Score.

Table 2. Six-month internet support group log-ins, posts, and process measures by engagement level across all groups.

Characteristic	Superusers (n=3)	Top contributors (n=30)	Contributors (n=108)	Observers (n=87)	Never log-ins (n=74)
ISG[a] log-ins					
Mean (SD)	119.0 (84.3)	22.5 (16.5)	7.3 (6.6)	2.5 (2.1)	N/A[b]
Median (IQR[c])	90 (53-214)	18 (13-27)	5.5 (3-9)	2 (1-3)	N/A
ISG posts					
Mean (SD)	210.0 (118.2)	17.2 (13.3)	3.2 (2.1)	N/A	N/A
Median (IQR)	246 (78-306)	11 (10-18)	3 (1-5)	N/A	N/A
cCBT[d] sessions completed					
Mean (SD)	8.0 (0.0)	7.4 (1.4)	5.8 (2.6)	4.2 (3.0)	1.9 (2.7)
Care manager contacts					
Mean (SD)	36.0 (11.8)	19.4 (5.6)	18.4 (6.4)	15.7 (5.0)	13.1 (4.7)

[a]ISG: internet support group.

[b]N/A: not applicable.

[c]IQR: interquartile range.

[d]cCBT: computerized cognitive behavioral therapy.

Table 3. Six-month internet support group log-ins, posts and process measures by engagement level with combined superusers + top contributors group.

Characteristic	Superusers + top contributors (n=33)	Contributors (n=108)	Observers (n=87)	Never log-ins (n=74)	*P* value
ISG[a] log-ins					
Mean (SD)	31.2 (38.5)	7.3 (6.6)	2.5 (2.1)	N/A[b]	<.001
Median (IQR[c])	20 (13, 31)	5.5 (3, 9)	2 (1, 3)	N/A	<.001[d]
ISG posts					
Mean (SD)	34.7 (64.8)	3.2 (2.1)	N/A	N/A	<.001
Median (IQR)	12 (10, 14)	3 (1, 5)	N/A	N/A	<.001[d]
cCBT[e] sessions completed					
Mean (SD)	7.4 (1.3)	5.8 (2.6)	4.2 (3.0)	1.9 (2.7)	<.001
Care manager contacts					
Mean (SD)	20.9 (7.8)	18.4 (6.4)	15.7 (5.0)	13.1 (4.7)	<.001

[a]ISG: internet support group.

[b]N/A: not applicable.

[c]IQR: interquartile range.

[d]*P* value from Kruskal-Wallis test.

[e]cCBT: computerized cognitive behavioral therapy.

Table 4. Mental health outcomes by engagement level across all groups[a].

Characteristic	Superusers + top contributors (n=33)	Contributors (n=108)	Observers (n=87)	Never log-ins (n=74)	*P* value
SF-12 MCS[b], estimated mean (SE)[c]					
Baseline	31.2 (1.9)	30.9 (1.2)	31.2 (1.2)	33.8 (1.3)	N/A[d]
6 months	47.2 (2.0)	42.5 (1.2)	43.8 (1.2)	44.0 (1.4)	N/A
Δ 6 months	16.1 (1.9)	11.7 (1.1)	12.6 (1.2)	10.1 (1.3)	.08
Superusers + top contributors vs never log-in	16.1 (1.9)	N/A	N/A	10.1 (1.3)	.01
PROMIS[e] Depression T-score, estimated mean (SE)[f]					
Baseline	62.0 (1.5)	62.3 (0.9)	61.8 (0.9)	61.1 (1.0)	N/A
6 months	51.7 (1.5)	54.1 (0.9)	53.2 (0.9)	53.5 (1.1)	N/A
Δ 6 months	–10.3 (1.3)	–8.2 (0.8)	–8.6 (0.8)	–7.6 (0.9)	.39
Superusers + top contributors vs never log-in	-10.3 (1.3)	N/A	N/A	-7.6 (0.9)	.09
PROMIS Anxiety T-score, estimated mean (SE)[g]					
Baseline	67.2 (1.5)	66.1 (0.9)	65.6 (0.9)	65.2 (1.0)	N/A
6 months	55.7 (1.5)	57.6 (0.9)	56.3 (0.9)	57.4 (1.1)	N/A
Δ 6 months	–11.6 (1.5)	–8.5 (0.8)	–9.4 (0.9)	–7.8 (1.0)	.19
Superusers + top contributors vs never log-in	-11.6 (1.5)	N/A	N/A	-7.8 (1.0)	.04

[a]All models are adjusted for gender, race, and education; n=259 (25 participants were missed at the 6-month assessment, and 9 participants withdrew from the study).

[b]SF-12 MCS: Short Form Health Survey Mental Components Score.

[c]Range 0-100; higher scores indicate better health-related quality of life.

[d]N/A: not applicable.

[e]PROMIS: Patient-Reported Outcomes Measurement Information System.

[f]T-score range 37.1-81.1; lower scores indicate less severe symptoms.

[g]T-score range 36.3-82.7; lower scores indicate less severe symptoms.

Table 5. Mental health outcomes by internet support group (ISG) log-in status[a].

Characteristic	Logged in to ISG ≥1 time[b] (n=228)	Never logged in to ISG (n=78)	*P* value
SF-12 MCS[c], estimated mean (SE)[d]			
Δ 6 months, log-in vs never log-in	12.6 (0.7)	10.1 (1.3)	.11
PROMIS[e]Depression T-score, estimated mean (SE)[f]			
Δ 6 months, log-in vs never log-in	-8.7 (0.5)	-7.6 (0.9)	.31
PROMIS Anxiety T-score, estimated mean (SE)[g]			
Δ 6 months, log-in vs never log-in	-9.3 (0.6)	-7.8 (1.0)	.20

[a]All models are adjusted for gender, race, and education; n=259 (25 participants were missed at the 6-month assessment, and 9 participants withdrew from the study).

[b]Includes superusers, top contributors, contributors, and observers.

[c]SF-12 MCS: Short Form Health Survey Mental Components Score.

[d]Range 0-100; higher scores indicate better health-related quality of life.

[e]PROMIS: Patient-Reported Outcomes Measurement Information System.

[f]T-score range 37.1-81.1; lower scores indicate less severe symptoms.

[g]T-score range 36.3-82.7; lower scores indicate less severe symptoms.

Discussion

Principal Findings

To the best of our knowledge, this is the first report to demonstrate that high levels of patient engagement with a moderated ISG, compared to no engagement with the ISG, are associated with improved anxiety symptoms and HRQoL in primary care. Our findings also provide further empirical evidence to support the participation inequality suggested by the 1% rule, although we observed a broader distribution of posting than posited by the 1% rule.

Our work confirms that depressed and anxious primary care patients are willing to engage in an ISG even when not required by study protocol to do so. Indeed, the sizable majority of our study subjects logged in to the ISG at least once, which is consistent with log-in rates reported in other studies of ISGs for depression [4,22]. Furthermore, among those who logged in to the ISG, participation inequality was less extreme than expected based on the 1% rule, as our top 1% and 10% of posters together generated 78% of all user-created content on our site, not 99% as the 1% rule predicts. Still, challenges remain in developing even more equitably engaged online communities to improve health and HRQoL.

Prior work on the impact of ISG engagement has been limited largely to comparing psychosocial outcomes between posters (defined as individuals who made at least 1 post) and observers in ISGs for women with breast cancer [23,24]. Findings from this work were mixed: while a moderate sized cross-sectional study showed more benefits in perceived social support in posters than observers [23], a large prospective study showed higher perceived functional well-being and fewer mood symptoms in observers than posters at 3-month follow-up [24]. To our knowledge, the only other study to explore the impact of engagement on mood symptoms in ISGs for mental health measured engagement by time spent on the ISG, showing that members who spent more than 5 hours on the ISG over a 2-week period were more likely to have resolution of depression at 6 months than members who spent less time [4].

Our finding that the participants who were highly engaged with the ISG reported improved anxiety symptoms and HRQoL at 6 months compared to individuals who never logged in identifies a subgroup that may benefit from participating in an ISG. Interestingly, this group did not report similar benefit for depression symptoms compared to the group that never logged in. On average, this highly engaged subgroup posted 5.8 times per month, which averaged approximately 1 post per log-in. Demographically, this subgroup had higher proportions of participants who were female, white, and college-educated than the group that never logged in, but both groups had similar levels of baseline depression and anxiety. This finding offers encouragement about the potential for ISGs to improve clinical outcomes in individuals who engage highly with an ISG. Still, more work is needed to confirm our findings in a randomized trial and identify the critical threshold of engagement needed to demonstrate clinically meaningful improvements in health.

Our work motivates further study into how to most accurately measure engagement with an ISG. We quantified engagement using the relatively simple metric of number of posts, and we assigned each post an equivalent weight. However, other quantitative metrics such as time spent on the ISG and number of pages viewed may offer a different perspective. Moreover, qualitative metrics that analyze post content may also be an important dimension of engagement, particularly considering evidence from breast cancer ISGs suggesting that a subset of members derive psychological benefit from creating posts that provide “insightful disclosure” [25].

Limitations

Our study has several limitations. First, our finding that high levels of engagement improved clinical outcomes reflects a post hoc analysis that we undertook to identify a subgroup that may have benefitted from the ISG. Second, the limited size of the ISG precluded further subgroup analyses and required us to combine the superuser and top contributor groups for all outcome analyses. Third, we quantified engagement using a simple measure of post counts, and we used this measure to stratify the sample into engagement levels based on the 1% rule rather than statistical methods that avoid specifying an a priori hypothesis about engagement distribution. Fourth, we were not able to include a content analysis of posts or examine the impact of post content on outcomes. Finally, since all participants had access to the cCBT program, we cannot exclude that the overall improvements we observed could be attributed to the cCBT program given its demonstrated efficacy [9,26].

Implications

Our work shows that primary care patient engagement in an online community for depression and anxiety may contribute to improved mental health at 6 months after enrollment but only at the highest levels of engagement. We strongly encourage researchers, clinicians, and health care delivery systems considering deployment of a similar ISG to first develop plans to encourage and sustain high and broad levels of user engagement. Future work is needed to (1) confirm our findings with mental health and other conditions, (2) establish the threshold of patient engagement required to benefit from an ISG, and (3) perform content and other qualitative analyses of discussion board posts to explore the influence of this content with patient engagement and other outcomes of interest.

Conclusions

In summary, we demonstrated that patient engagement with our moderated ISG for depressed and anxious primary care patients generally approximates the 1% rule. Although we observed broader engagement levels with the ISG than predicted by the 1% rule, only ISG members who engaged at the highest levels of engagement reported measurable improvement in symptoms and quality of life at 6-month follow-up.

Acknowledgments

This work was supported by grant R01 MH093501 from the National Institute of Mental Health and grant TL1 TR000145 from the National Center for Advancing Translational Sciences of the National Institutes of Health.

Authors' Contributions

EMG, BHB, KZA, and BLR designed the study. KZA and SDR analyzed the data. EMG, BHB, KZA, SDR, AJR, and BLR drafted the manuscript. All authors critically revised and edited the draft and approved of the final version. EMG, BHB, KZA, SDR, AJR, and BLR had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of our internet support group.

[PPTX File, 1MB - jmir_v20i7e10402_app1.pptx]

References

1. van der Eijk M, Faber MJ, Aarts JWM, Kremer JAM, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013 Jun;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
2. Rollman BL, Herbeck BB, Rotondi AJ. Internet support groups for health: ready for the Affordable Care Act. J Gen Intern Med 2014 Nov;29(11):1436-1438 [FREE Full text] [doi: 10.1007/s11606-014-2884-z] [Medline: 24817281]
3. Callan JA, Wright J, Siegle GJ, Howland RH, Kepler BB. Use of computer and mobile technologies in the treatment of depression. Arch Psychiatr Nurs 2017 Dec;31(3):311-318. [doi: 10.1016/j.apnu.2016.10.002] [Medline: 28499574]
4. Houston TK, Cooper LA, Ford DE. Internet support groups for depression: a 1-year prospective cohort study. Am J Psychiatry 2002 Dec;159(12):2062-2068. [Medline: 12450957]
5. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/appi.ps.61.11.1099] [Medline: 21041348]
6. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
7. Griffiths KM, Calear AL, Banfield M. Systematic review on Internet Support Groups (ISGs) and depression (1). Do ISGs reduce depressive symptoms? J Med Internet Res 2009;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]
8. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
9. Rollman BL, Herbeck BB, Abebe KZ, Spring MB, Rotondi AJ, Rothenberger SD, et al. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: a randomized clinical trial. JAMA Psychiatry 2018 Jan 01;75(1):56-64. [doi: 10.1001/jamapsychiatry.2017.3379] [Medline: 29117275]
10. Arthur C. What is the 1% rule?.: The Guardian; 2006. URL: https://www.theguardian.com/technology/2006/jul/20/guardianweeklytechnologysection2 [accessed 2018-06-12] [WebCite Cache ID 708CXPPA5]
11. Nielsen J. The 90-9-1 rule for participation inequality in social media and online communities. 2006. URL: https://www.nngroup.com/articles/participation-inequality/ [accessed 2018-06-12] [WebCite Cache ID 708CgyqsV]
12. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
13. van Mierlo T, Li X, Hyatt D, Ching AT. Demographic and indication-specific characteristics have limited association with social network engagement: evidence from 24,954 members of four health care support groups. J Med Internet Res 2017 Feb 17;19(2):e40 [FREE Full text] [doi: 10.2196/jmir.6330] [Medline: 28213340]
14. 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. [doi: 10.1016/j.invent.2017.02.003] [Medline: 28584734]
15. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
17. Proudfoot J, Ryden C, Everitt B, Shapiro DA, Goldberg D, Mann A, et al. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:46-54 [FREE Full text] [Medline: 15231555]
18. Rollman BL, Herbeck BB, Reynolds CF, Schulberg HC, Shear MK. A contemporary protocol to assist primary care physicians in the treatment of panic and generalized anxiety disorders. Gen Hosp Psychiatry 2003;25(2):74-82. [Medline: 12676419]
19. Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. JAMA 1994 Dec 14;272(22):1749-1756. [Medline: 7966923]
20. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [Medline: 8628042]
21. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, PROMIS Cooperative Group. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): depression, anxiety, and anger. Assessment 2011 Sep;18(3):263-283 [FREE Full text] [doi: 10.1177/1073191111411667] [Medline: 21697139]
22. Griffiths KM, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, Farrer L. The effectiveness of an online support group for members of the community with depression: a randomised controlled trial. PLoS One 2012;7(12):e53244 [FREE Full text] [doi: 10.1371/journal.pone.0053244] [Medline: 23285271]
23. Setoyama Y, Yamazaki Y, Namayama K. Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. J Med Internet Res 2011;13(4):e122 [FREE Full text] [doi: 10.2196/jmir.1696] [Medline: 22204869]
24. Han JY, Hou J, Kim E, Gustafson DH. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. Health Commun 2014;29(9):911-923 [FREE Full text] [doi: 10.1080/10410236.2013.816911] [Medline: 24345206]
25. Shaw BR, Hawkins R, McTavish F, Pingree S, Gustafson DH. Effects of insightful disclosure within computer mediated support groups on women with breast cancer. Health Commun 2006;19(2):133-142. [doi: 10.1207/s15327027hc1902_5] [Medline: 16548704]
26. Cavanagh K, Herbeck Belnap B, Rothenberger S, Abebe K, Rollman B. My care manager, my computer therapy and me: the relationship triangle in computerized cognitive behavioural therapy. Internet Interventions 2018 Mar;11:11-19. [doi: 10.1016/j.invent.2017.10.005]

Abbreviations

cCBT: computerized cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder scale
HRQoL: health-related quality of life
IQR: interquartile range
ISG: internet support group
PCP: primary care physician
PHQ-9: Patient Health Questionnaire
PRIME-MD: Primary Care Evaluation of Mental Disorders
PROMIS: Patient-Reported Outcomes Measurement Information System
SF-12 MCS: Short Form Health Survey Mental Components Score

Edited by G Eysenbach; submitted 14.03.18; peer-reviewed by A Kapoor, G Strudwick; comments to author 26.04.18; revised version received 14.05.18; accepted 19.05.18; published 17.07.18.

Please cite as:
Geramita EM, Herbeck Belnap B, Abebe KZ, Rothenberger SD, Rotondi AJ, Rollman BL
The Association Between Increased Levels of Patient Engagement With an Internet Support Group and Improved Mental Health Outcomes at 6-Month Follow-Up: Post-Hoc Analyses From a Randomized Controlled Trial
J Med Internet Res 2018;20(7):e10402
URL: http://www.jmir.org/2018/7/e10402/
doi:10.2196/10402
PMID:30021711

Original Paper

Privacy, Trust, and Data Sharing in Web-Based and Mobile Research: Participant Perspectives in a Large Nationwide Sample of Men Who Have Sex With Men in the United States

H Jonathon Rendina[1,2], PhD, MPH; Brian Mustanski[3], PhD

[1]Department of Psychology, Hunter College, The City University of New York, New York, NY, United States

[2]Doctoral Program in Health Psychology and Clinical Science, The Graduate Center, The City University of New York, New York, NY, United States

[3]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
H Jonathon Rendina, PhD, MPH
Department of Psychology
Hunter College
The City University of New York
695 Park Avenue
New York, NY, 10065
United States
Phone: 1 2122067919 ext 922
Email: hrendina@hunter.cuny.edu

Abstract

Background: Modern research is heavily reliant on online and mobile technologies, which is particularly true among historically hard-to-reach populations such as gay, bisexual, and other men who have sex with men (GBMSM). Despite this, very little empirical research has been published on participant perspectives about issues such as privacy, trust, and data sharing.

Objective: The objective of our study was to analyze data from an online sample of 11,032 GBMSM in the United States to examine their trust in and perspectives on privacy and data sharing within online and mobile research.

Methods: Participants were recruited via a social networking site or sexual networking app to complete an anonymous online survey. We conducted a series of repeated measures analyses adjusted for between-person factors to examine within-person differences in the following: (1) trust for guarding personal information across different venues (eg, online research conducted by a university vs. an online search engine); (2) privacy concerns about 12 different types of data for three distinct data activities (ie, collection by app owners, anonymous selling to third parties, and anonymous sharing with researchers); and (3) willingness to share those 12 different types of data with researchers. Due to the large sample size, we primarily reported measures of effect size as evidence of clinical significance.

Results: Online research was rated as most trusted and was more trusted than online and mobile technology companies, such as app owners and search engines, by magnitudes of effect that were moderate-to-large ($\eta_{partial}^2$=0.06-0.11). Responding about 12 different types of data, participants expressed more concerns about data being anonymously sold to third-party partners (mean 7.6, median 10.0) and fewer concerns about data being collected by the app owners (mean 5.8, median 5.0) or shared anonymously with researchers (mean 4.6, median 3.0); differences were small-to-moderate in size ($\eta_{partial}^2$=0.01-0.03). Furthermore, participants were most willing to share their public profile information (eg, age) with researchers but least willing to share device usage information (eg, other apps installed); the comparisons were small-to-moderate in size ($\eta_{partial}^2$=0.03).

Conclusions: Participants reported high levels of trust in online and mobile research, which is noteworthy given recent high-profile cases of corporate and government data security breaches and privacy violations. Researchers and ethical boards should keep up with technological shifts to maintain the ability to guard privacy and confidentiality and maintain trust. There was substantial variability in privacy concerns about and willingness to share different types of data, suggesting the need to gain consent for data sharing on a specific rather than broad basis. Finally, we saw evidence of a privacy paradox, whereby participants expressed privacy concerns about the very types of data-related activities they have likely already permitted through the terms of the apps and sites they use regularly.

(J Med Internet Res 2018;20(7):e233) doi:10.2196/jmir.9019

KEYWORDS

data privacy; data sharing; research trust; mobile research; research ethics; men who have sex with men; gay and bisexual men

Introduction

Since the development of the "World Wide Web" nearly three decades ago, the diversity and usage of available online and mobile technologies have proliferated rapidly, resulting in a shift in the landscape of their use among various populations. Among gay, bisexual, and other men who have sex with men (GBMSM), these shifts have been evident in the use of these technologies for sexual networking, which has developed from online computer chat rooms to mobile geosocial networking applications (ie, "apps") to identify potential partners by various characteristics and categorize them by distance [1]. Evidence suggests that a large number of GBMSM use these technologies to locate sexual partners [2]. As a result of their popularity, researchers have leveraged these technologies to reach and recruit GBMSM—who have historically been a hidden or hard-to-reach population [3]—into formative and intervention studies, particularly on a range of HIV prevention and treatment topics [4-14]. Although most research, both with and without such technologies, has traditionally focused on HIV, the focus is gradually broadening [15], given that GBMSM are part of the broader population of sexual and gender minorities and are now recognized by the US National Institutes of Health as a "health disparity population" [16]. As the GBMSM-focused research agenda broadens, it is also likely to shift more toward online and mobile methodologies because of their popularity. However, this surge in research using technology brings with it novel methodological and ethical considerations.

When making decisions regarding the ethical implications of online and mobile research, researchers and review boards are charged with evaluating and minimizing risk to participants, but rapidly evolving technological advances have made it difficult to keep pace [17,18]. Researchers and ethical review boards experience several issues related to human subjects' protections in online and mobile research that are either unique or different from those encountered in traditional research, including issues of informed consent, privacy/confidentiality, data security, and ownership of and access to data [19-25]. In recent years, there has been a substantial increase in the number of scientific papers reviewing the state of science, empirically evaluating or discussing the implications of privacy, security, and confidentiality in online and mobile research [17,19,20,22-29]. Besides risks that offline and online research share, one of the primary forms of risk posed by online and mobile research is that of informational risk [20,30], which is risk that research might lead to unintended creation, tracking, or sharing of data with third parties or interception of data by other audiences [23]. For example, many individuals—whether potential participants, researchers themselves, or ethics reviewers—are unaware of the extent to which third-party marketing firms could track and store information about individuals' internet behaviors (eg, clicking on an ad for a research study) to create a complex profile of individuals for advertising purposes [20]. To the extent that such data, even if minimally detailed, are collected by app owners without the knowledge of researchers or participants, issues can arise about understanding and protecting the privacy and confidentiality of participants, thus potentiating research mistrust.

In addition to understanding the technical and legal aspects of risks when using online and mobile technologies, it is important to understand and weigh participants' perspectives on trust, privacy, and data sharing. Regarding issues of privacy and confidentiality when using online and mobile technology for personal rather than research purposes, views continue to develop among the general public together with the changing technological landscape [31-35]. For example, data on 461 adults in the United States collected by Pew Research Center suggested that people weigh tradeoffs between disclosing personal information and the benefits of doing so; more than half (56.8%, 262/461) considered it acceptable to use a health information website that their doctor would upload their health data to as long as it was secure (ie, high benefits and low likelihood of disclosures), whereas only one-third (33.2%, 153/461) considered it acceptable to use a social media site that would use their profile data to deliver targeted advertising [36]. Subsequent Pew data highlighted the general public's trust in online and mobile companies that they regularly use; data from 1040 US adults in 2016 suggested that 65.3% (605/926) were somewhat or very confident that their email providers adequately safeguard the privacy of their data, although this figure was only 47.2% (314/665) for social media sites that they used [37].

Compared with the available data on participants' perspectives on privacy within the technologies they use for personal reasons, fewer published studies are available regarding participants' perspectives on these issues in online and mobile research. Nonetheless, the available data suggest equally nuanced and developing views. One study showed that people preferred online methods over traditional means of research and considered online research to be *more* private than traditional in-person methods, although submitting sensitive health and personal information emerged as a concern [38]. Another recent study reported that concerns about privacy and confidentiality in online and mobile research are diverse and often contextually specific, varying across individuals, as well as by the type of data, the context of data collection, and the purpose of data collection or usage [39,40]. Echoing the findings outside the research context, these studies suggested that participants value control over whether and how data are used.

Although the data above highlight participants' viewpoints regarding privacy in online and mobile research from the general public and despite growing literature on methodological issues related to online and mobile research with GBMSM [1,4,6,8,9], we are unaware of any published research that has examined the perceptions of privacy in online and mobile research specific to GBMSM. Given the growing use of online and mobile technologies within research with GBMSM and the relatively

unique technologies available to and used by GBMSM, it is imperative to understand their views about the risks and benefits of technology-based research. We believe such data will be of use to future researchers as they design technology-based studies, consider industry partnerships to conduct research, and weigh the risks and benefits of such designs.

This study was designed to fill the noted gaps in the literature on GBMSM perspectives on trust, privacy, and data sharing in online and mobile research and to achieve three aims. First, we sought to understand trust in online and mobile research compared with that in the use of online and mobile technologies for everyday purposes. Thus, we compared levels of trust for guarding personal information—defined broadly—across numerous sources that collect such data (eg, an online research study vs. a social networking website). Second, we sought to better understand which specific types of data caused participants more and less concern about privacy. We compared the extent of privacy concerns endorsed for three distinct practices within a hypothetical app—collection and storage of the data by app developers, sale of data anonymously to third-party partners, and sharing of data anonymously with researchers—across a range of unique types of personal data. Third, we sought to examine willingness to have different types of app-generated data shared with researchers. Using the same unique types of personal data from the second aim, we compared hypothetical willingness to provide consent to have an app developer/owner share these different types of data anonymously with researchers.

Methods

In this study, data were reported from an extensive nationwide survey of GBMSM conducted over a 4-week period between May and June 2017.

Participants and Procedures

Between May 17, 2017 and June 10, 2017, we used advertisements to enroll GBMSM from two venues—one of the most popular geotargeted sexual networking apps for GBMSM and one of the most popular social networking websites for the general population. The sexual networking app pushed the advertisement as a message to the chat inboxes of all users in the United States on Friday, May 19, 2017, which remained for 7 days, unless deleted sooner. On the social networking site, we used targeted banner advertisements for approximately 4 weeks that could show up in one of the two ways—a static ad on the right-hand pane of the website or an ad that resembled a normal post as users scrolled through their feeds. We targeted the social networking site ads to people who were men, residing in the United States, aged ≥18 years, and believed to be GBMSM based on either a same-sex interest listed on their profile or a range of relevant "likes" (eg, gay pride, lesbian, gay, bisexual, and transgender, LGBT, community, gay bar, and same-sex marriage). Both ads comprised a background image (the social networking site: 2 clothed men on a bed kissing; the sexual networking app, 2 bare torsos embracing) and brief text, including that they could "enter to win a $50 Amazon.com gift card" and that there was "no participation necessary" to enter the random drawing.

Upon clicking on the ad, the participants were informed that the survey would take approximately 10-15 min to complete and provided the option to begin immediately or enter their email address to receive a link to complete later. Upon beginning the survey—whether immediately or through the emailed link—participants were provided with a brief online consent form and given the options of providing consent, declining, or declining with the option to receive instructions for entering the random gift card drawing. During consent, participants were informed about a 1 in 100 chance of receiving a $50 gift card. Those who subsequently declined consent were provided instructions should they want to enter the drawing; conversely, participants who completed the survey and were interested in entering the drawing were redirected to a separate survey in which they were required to enter an email address that was not linked in any way to their data. During the first few questions of the survey, participants were screened for eligibility, which was defined as follows: (1) age ≥18 years; (2) residing in the United States; (3) having had same-sex sexual activity within the past year; and (4) identifying as male (including both cisgender and transgender males). Those who were ineligible were informed after the first few questions, and the survey subsequently ended. The study protocol was reviewed and approved by the Human Research Protections Program of The City University of New York (New York, NY, USA).

We followed a protocol based on standards within the literature [41] for removing potentially duplicate cases while erring on the side of keeping rather than removing data in cases where a determination could not be made. In particular, we first identified potential duplicates based on birth month and year, zip code, HIV status, and race/ethnicity; all cases sharing those features in common were manually examined, focusing on responses to other questions such as education, employment, and partner status, as well as device and browser information and the survey duration.

Measures

We collected all measures for this study as self-reported items and scales within the one-time online survey. The item content was developed in part by consulting the terms of service and privacy policy for two social networking (ie, Facebook and Facebook Messenger) and two sexual networking (ie, Grindr and Scruff) apps in late 2016. In addition, we examined the types of personal information and data discussed within those agreements and the usage provided for within the agreements to develop three primary data activities described in the measure below (ie, data collection, anonymous sale of data, and anonymous sharing of data). Likewise, we used the sites and apps to create a list of the types of personal information (ie, data) that are likely to be gathered and/or generated by developers. After obtaining the complete draft of the measures, we invited a group of 20 adult GBMSM in the New York City area to participate in an in-person community feedback session; all participants were provided with a copy of the measures, and we reviewed both the study procedures (eg, recruitment and compensation) and the item content with them to gather their feedback. We received and followed numerous suggestions to improve clarity, reduce length, and minimize burden. For example, from a list of at least 15 different types of data,

community members noted that they were not all meaningfully distinct; thus, the list was condensed to form broader categories in some cases. Similarly, we implemented suggestions for improved wording. The final version of the measures was based on this feedback and a review by field experts from Fordham University's HIV Prevention and Substance Use Research Ethics Training Institute (New York, NY), as described later (see the Online Supplementary Material for more details).

Demographic Characteristics

Participants responded to items inquiring about various demographic characteristics, including age, zip code (which was converted to geographic region), relationship status, sexual orientation, and race/ethnicity.

Trust to Safeguard Personal Information

All participants received the following instructions:

"We are interested in knowing more about how much you trust various organizations and businesses to protect the privacy and confidentiality of the data they collect on you. Please assume you are being asked to provide similar information to each. How much do you trust that each of the following sources would guard the privacy and confidentiality of your personal information?"

Following this, they were presented with a list of nine different types of online and mobile venues in which personal information could be collected and asked to rate their trust on a scale from 1 (*Not at all trusting*) to 4 (*Very trusting*).

Concerns About Privacy Threats

We presented the participants with a vignette describing a hypothetical new app with various features. Then, a series of 12 types of personal information were presented and participants were asked, for each, whether the following activities concerned them as a threat to their privacy: (1) app owners privately collecting and storing these data; (2) app owners selling these data anonymously to third-party marketing groups; and (3) app owners sharing these data anonymously with researchers. Participants were asked to check which, if any, of the three activities concerned them separately for each of the 12 types of personal information (ie, a total of 36 dichotomous responses).

Data Sharing With Researchers

Finally, we presented the participants with the same 12 types of personal information from the prior measure and the following instructions:

"Within this study, we are not gathering any data on you from any apps or sites that you use. However, please imagine we were interested in connecting data collected by the app with the data you provided in this survey. Which of the following would you give us permission to gather anonymously from the app owners to link with your survey data?"

Participants rated their willingness to provide permission for each on a scale from 1 (*Definitely not*) to 4 (*Definitely*).

Data Analysis

All analyses were performed in SPSS 24 (IBM Corporation; Amonk, New York, United States). To inform future online recruitment efforts, we began our analyses by characterizing the sociodemographic characteristics of the sample and comparing them across the two recruitment venues using chi-square tests of independence. To address the first aim regarding the comparisons of trust for guarding personal information across nine different sources, we iteratively conducted a series of 36 repeated measures analysis of variance (RMANOVA) models examining each pair of ratings while adjusting for relevant between-person characteristics (ie, recruitment source, race, HIV status, and age); we specified an interaction for each between-person factor with the within-person factor but not among the between-person factors. We reported the $\eta_{partial}^2$ effect sizes for the within-person main effect as evidence of the magnitude of each comparison. To address the second aim regarding privacy concerns raised about 12 different types of app-related data across 3 different data activities (ie, the app collecting the data, the app anonymously selling the data, and the app anonymously sharing the data with researchers), we assessed the prevalence of indicating each was a concern by examining the frequency and proportion of "yes" responses across the 36 dichotomous indicators. We also calculated a sum score for the total number of types of data that raised concerns for participants for each of the 3 data activities and compared the 3 sum scores to one another in an RMANOVA that was consistent with the prior set of analyses with two exceptions—all 3 scores were compared simultaneously rather than in pairs, and we used a simple contrast to test differences between the three, using sharing with researchers as the referent group. Finally, to address the third aim regarding which types of app-related data participants would hypothetically be willing to provide explicit permission to have shared with researchers, we used the same 12 types of data asked about in the second aim and used a series of 66 pairwise RMANOVAs consistent with the first set of analyses to compare within-person differences among the 12 ratings adjusted for the relevant between-person factors.

Across all analyses, the primary goal was to examine patterns in the data descriptively using effect sizes rather than search for statistical significance, particularly because of the large sample size. Furthermore, we reported the $\eta_{partial}^2$ effect size as small (0.01), medium/moderate (0.06), and large (0.14) in size [42]. Nonetheless, for statistical comparisons, we reported statistical significance for those findings that reached a threshold of $P<.001$ to reduce the likelihood of type II error because of multiple comparisons.

We conducted an experimental manipulation to test whether providing a rationale for each of the 3 activities within the "*Concerns about privacy threats* " measure would influence trust. Specifically, participants were randomized to receive either a description of the 3 activities with no rationale or the same description with rationale added (eg, for the app owners collecting the information, rationale added was "to improve, tailor, and develop the services you use"). As results suggested nonsignificant and extremely small (Cohen $d<.05$) differences between groups, all results are presented irrespective of the experimental condition.

Results

Upon reaching the landing page of the survey from the advertisement, 80.4% (21,942/27,291) of participants agreed to be immediately linked to the survey, 17.1% (4677/27,291) opted to receive an email and complete the survey at a later time, and 2.5% (672/27,291) opted not to take the survey. Subsequently, 18,909 reached the consent form, of whom 94.9% (17,954/18,909) provided consent, 1.4% (262/18,909) declined consent, and 3.7% (693/18,909) requested instructions on how to enter the drawing without completing the survey. Of 17,954 who provided consent, 7.4% (1335/17,954) did not provide sufficient data to determine eligibility, 11.5% (2068/17,954) were deemed ineligible, 19.4% (3487/17,954) were eligible but only partially completed the survey, and 61.6% (11,064/17,954) completed the survey in its entirety. Among those who reached the consent form, the completion rates were similar for those who began the survey from the social networking site (56.6%, 2193/3874) and the sexual networking app (59.0%, 8871/15,035). Finally, of the completed surveys, we eliminated 30 completed surveys that were duplicate responses of previously completed surveys, resulting in a final analytic sample of 11,032 GBMSM in the United States.

Table 1 summarizes the sociodemographic characteristics of the analytic sample with comparisons by recruitment source—nearly one-fifth (19.6%, 2166/11,032) were recruited from the social networking site and the remainder (80.4%, 8866/11,032) were enrolled from the sexual networking app. The sample was diverse regarding race/ethnicity, with nearly half (46.2%, 5102/11,032) being men of color. Most of the sample identified as cisgender male (98.5%, 10,869/11,032), and gay or queer (81.9%, 9045/11,032) and the majority reported being HIV-negative (75.0%, 8275/11,032); we observed diversity in employment, educational experiences, and geographic regions. In addition, we observed significant differences between the 2 recruitment sources regarding race/ethnicity, gender identity, sexual identity, employment status, and geographic region; the sexual networking app comprised more men of color, fewer transgender males, more nongay identified men, more men who were working full-time, and fewer men from the South. The sample ranged in the 18-80 years of age, with an average age of 32.6 (SD 12.0; median 29.0) years, with the social networking site (mean 33.3, SD 14.3) being 1 year older, on average, than the sexual networking app (mean 32.4, SD 11.3).

Table 2 presents the ratings of trust to guard personal information by source, with corresponding within-person comparisons across all sources reported as $\eta_{partial}^2$ effect sizes. As evident within the unadjusted means, participants rated the 3 types of online research studies with a high degree of trust for guarding personal information—the median rating for each was a 3 on a range of 1-4 with minimal differences between them. The next most trusted source was the partnership between researchers and a mobile app for GBMSM, which exhibited minimal differences in trust ratings from those of the 3 types of online research and medium-to-large differences from each of the 5 types of online and mobile companies. The mobile app designed for GBMSM was rated much lower than the four types of online research and slightly higher than the online and mobile technologies for the general public based on the unadjusted means; however, the adjusted within-person comparisons revealed inconsequentially small differences in rating between the GBMSM-specific app and each of the 3 types of online and mobile companies for the general public, which also had minimal differences from one another.

Table 3 presents the prevalence of data concerns by each type of data and data activity (ie, data collection, anonymous data sale to third parties, and anonymous data sharing with researchers). Here, two trends are worth noting. First, across the 3 activities, there was diversity in terms of which *types* of data participants were concerned about—the most widely endorsed types of data that concerned participants were device data, such as global positioning system (GPS) information and information about other apps installed on the phone, whereas the least endorsed were about usage of the app such as how often one logged in or whether they participated in any app-based health promotion campaigns. Second, regardless of the *type* of data, there was a trend about the data *activities* that were the most concerning, with a marked number of participants endorsing a concern about the anonymous sale of their data to third-party partners and the lowest numbers endorsing concern about the anonymous sharing of their data with researchers. Notably, across each type of data, more participants expressed concern about the app collecting their data in the first place than did about the anonymous sharing of their data with researchers. Table 3 also presents the average number of types of data endorsed as concern for each of the 3 activities. In within-person comparisons, we found that all three were significantly different from one another ($\eta_{partial}^2$=0.01; P<.001), with a small-to-moderate difference between anonymous data sharing with researchers and anonymous data selling to third parties ($\eta_{partial}^2$=0.03) and a small difference between sharing with researchers and collection of the data themselves ($\eta_{partial}^2$=0.01).

Table 1. Sociodemographic characteristics and comparisons by the recruitment source.

Characteristics	Full sample (N=11,032), n (%)	Social networking site (n=2166), n (%)	Sexual networking app (n=8866), n (%)	χ^2 (*df*)
Race/Ethnicity				216.4 (4)[a]
Black	1100 (10.0)	107 (4.9)	993 (11.2)	
Latino	2409 (21.8)	344 (15.9)	2065 (23.3)	
White	5930 (53.8)	1456 (67.2)	4474 (50.5)	
Multiracial	808 (7.3)	142 (6.6)	666 (7.5)	
Other	785 (7.1)	117 (5.4)	668 (7.5)	
Gender Identity				198.9 (1)[a]
Cisgender male	10869 (98.5)	2063 (95.2)	8806 (99.3)	
Transgender male	163 (1.5)	103 (4.80)	60 (0.7)	
Sexual Identity				33.8 (3)[a]
Gay, queer, or homosexual	9045 (82.0)	1862 (86.0)	7183 (81.0)	
Bisexual	1802 (16.3)	275 (12.7)	1527 (17.2)	
Heterosexual	46 (0.4)	2 (0.1)	44 (0.5)	
Other	139 (1.3)	27 (1.2)	112 (1.3)	
Employment Status				57.6 (3)[a]
Full-time	5990 (54.3)	1038 (47.9)	4952 (55.9)	
Part-time	2505 (22.7)	528 (24.4)	1977 (22.3)	
On disability	655 (5.9)	180 (8.3)	475 (5.4)	
Unemployed	1882 (17.1)	420 (19.4)	1462 (16.5	
Educational Attainment				0.6 (3)
High school, GED[b], or less	2395 (21.7)	469 (21.7)	1926 (21.7)	
Some college	4908 (44.5)	975 (45.0)	3933 (44.4)	
4-year college degree	2434 (22.1)	465 (21.5)	1969 (22.2)	
Postgraduate degree	1295 (11.7)	257 (11.9)	1038 (11.7)	
HIV Status				9.0 (2)
Negative	8275 (75.0)	1679 (77.5)	6596 (74.4)	
Positive	1837 (16.7)	326 (15.1)	1511 (17.0)	
Unknown	920 (8.3)	161 (7.4)	759 (8.6)	
Geographic Region				20.3 (4)[a]
Northeast	2089 (18.9)	400 (18.5)	1689 (19.1)	
South	2045 (18.5)	469 (21.7)	1576 (17.8)	
Midwest	3777 (34.2)	699 (32.3)	3078 (34.7)	
West	3034 (27.5)	587 (27.1)	2447 (27.6)	
Other/Unknown	87 (0.8)	11 (0.5	76 (0.9)	

[a]*P*<.001

[b]GED: General Equivalency Diploma.

Table 2. Within-person comparisons of trust to guard the privacy of personal information reported as $\eta_{partial}^2$ effect sizes. Results are reported as $\eta_{partial}^2$ effect sizes for the difference between the two means adjusted for demographic covariates (eg, unadjusted means, medians, and standard deviations are presented in the far right columns to ease interpretation of the comparisons). Response options ranged from 1 (*not at all trusting*) to 4 (*very trusting*).

Source	1	2	3	4	5	6	7	8	9	Mean (SD)	Median
1. Online research study by researchers at a university	—									2.82 (0.84)	3.00
2. Online research study by an LGBT[a] community center	0.00	—								2.87 (0.82)	3.00
3. Online research study by government health agency	0.00[b]	0.00[b]	—							2.81 (0.96)	3.00
4. Mobile networking app for GBMSM[c]	0.08[b]	0.10[b]	0.09[b]	—						2.03 (0.84)	2.00
5. Mobile networking app for the general public	0.10[b]	0.11[b]	0.11[b]	0.00[b]	—					1.81 (0.81)	2.00
6. Online shopping website	0.06[b]	0.07[b]	0.08[b]	0.00	0.00	—				1.84 (0.91)	2.00
7. Online email website	0.06[b]	0.06[b]	0.07[b]	0.00	0.00[b]	0.00	—			1.83 (0.90)	2.00
8. Online search engine	0.06[b]	0.07[b]	0.08[b]	0.00	0.00[b]	0.00	0.00	—		1.83 (0.90)	2.00
9. Research study by researchers at a university in collaboration with mobile networking app for GBMSM	0.00	0.00	0.01[b]	0.09[b]	0.10[b]	0.06[b]	0.06[b]	0.06[b]	—	2.69 (0.87)	3.00

[a]LGBT: lesbian, gay, bisexual, and transgender.

[b]$P<.001$.

[c]GBMSM: gay, bisexual, and other men who have sex with men.

Table 3. Prevalence of privacy concerns by type of data and data activity. Numbers and percentages correspond to those participants who endorsed each item as a concern.

Type of data	App owners collecting, n (%)	App owners anonymously selling to third parties, n (%)	App owners anonymously sharing with researchers, n (%)
Public profile information (eg, age and height)	5523 (50.1)	7302 (66.2)	4081 (37.0)
Account information (eg, birthdate and zip code)	5418 (49.1)	7500 (68.0)	4106 (37.2)
Match information (eg, HIV status and dating interests)	5039 (45.7)	7016 (63.9)	4090 (37.1)
Mobile device information (eg, operating system)	5187 (47.0)	6901 (62.6)	4258 (38.6)
Interaction information (eg, demographics of chat partners)	5251 (47.6)	6964 (63.1)	4143 (37.6)
App usage information (eg, login frequency)	4843 (43.9)	6459 (58.5)	3783 (34.3)
Health campaign participation information (eg, HIV test reminders)	4713 (42.7)	6540 (59.3)	3867 (35.1)
Device GPS[a] information (eg, login locations)	6020 (54.6)	7469 (67.7)	4735 (42.9)
Device usage information (eg, other apps installed)	6337 (57.4)	7563 (68.6)	5138 (46.6)
App advertising information (eg, ad clicks)	5047 (45.7)	6890 (62.5)	4123 (37.4
Third-party advertiser information (eg, service utilization)	5165 (46.8)	6880 (62.4)	4143 (37.6
App-generated information (eg, advertising profiles)	5016 (45.5)	6810 (61.7)	4043 (36.6)
Total number of concerns (range: 0-12), mean (median)	5.8 (5.0)	7.6 (10.0)	4.6 (3.0)

Table 4. Willingness to share various data types with researchers and within-person comparisons between each reported as $\eta_{partial}^2$ effect sizes. Results are reported as $\eta_{partial}^2$ effect sizes for the difference between the two means adjusted for demographic covariates (eg, unadjusted means, medians, and standard deviations are presented in the far right columns to ease interpretation of the comparisons). Responses ranged from 1 (*definitely not*) to 4 (*definitely*).

Type of data	1	2	3	4	5	6	7	8	9	10	11	12	Mean (SD)	Median
1. Public profile information	—												2.85 (0.93)	3.00
2. Account information	0.05[a]	—											2.35 (1.00)	2.00
3. Match information	0.01[a]	0.02[a]	—										2.65 (0.97)	3.00
4. Mobile device information	0.03[a]	0.00	0.01[a]	—									2.27 (1.03)	2.00
5. Interaction information	0.02[a]	0.01[a]	0.01[a]	0.00[a]	—								2.35 (0.99)	2.00
6. App usage information	0.01[a]	0.02[a]	0.00	0.01[a]	0.01[a]	—							2.48 (0.98)	3.00
7. Health campaign participation	0.00[a]	0.02[a]	0.00	0.02[a]	0.01[a]	0.00	—						2.58 (0.98)	3.00
8. Device GPS[b] information	0.06[a]	0.01[a]	0.04[a]	0.01[a]	0.02[a]	0.04[a]	0.05[a]	—					2.08 (1.02)	2.00
9. Device usage information	0.06[a]	0.01[a]	0.04[a]	0.01[a]	0.02[a]	0.04[a]	0.05[a]	0.00	—				1.92 (1.00)	2.00
10. App advertising information	0.03[a]	0.00[a]	0.01[a]	0.00	0.00	0.01[a]	0.01[a]	0.02[a]	0.02[a]	—			2.24 (0.99)	2.00
11. Third-party advertiser information	0.03[a]	0.00[a]	0.01[a]	0.00	0.00[a]	0.01[a]	0.02[a]	0.01[a]	0.02[a]	0.00	—		2.21 (0.99)	2.00
12. App-generated information	0.02[a]	0.01[a]	0.00[a]	0.00[a]	0.00	0.00[a]	0.01[a]	0.03[a]	0.03[a]	0.00[a]	0.01[a]	—	2.33 (0.99)	2.00

[c]$P<.001$.

[a]GPS: global positioning system.

Besides knowing which types of data collection, sale, and sharing are of concern as a threat to participant's privacy, we were also interested in determining which types of data they would give explicit permission to researchers to request from app owners. Table 4 presents the average willingness expressed by participants for each type of data, which were similar to those examined in the prior set of analyses. We observed a range of willingness across the 12 types of data with adjusted within-person differences between different types of data ranging from very small ($\eta_{partial}^2<0.01$) to medium ($\eta_{partial}^2=0.06$); the majority of participants were willing to share 4 types of data and unwilling to share the other 8 (ie, 4 had median values of 3.0 corresponding to probable willingness). Consistent with the previous aim's findings on which types of data represented a privacy concern, participants were least willing to share those data that were generated by their devices such as GPS and other apps installed, whereas they were most willing to share general app information such as public profile and match survey data, as well as app usage statistics and health campaign participation. In fact, the largest differences were in comparing the public profile information with device GPS data ($\eta_{partial}^2=0.06$) and device usage information ($\eta_{partial}^2=0.06$).

Discussion

Primary Findings

We analyzed data from an online sample of 11,032 GBMSM across the United States to examine participant perspectives on the issues of trust, privacy, and data sharing in online and mobile research. In analyses that were adjusted for relevant between-person differences (including the recruitment site), we found that trust in online research was greater than trust in online and mobile platforms for personal use, such as social and sexual networking apps or various types of websites. When focusing on 12 different types of data that could be gathered by a hypothetical sexual networking app, participants expressed the least concerns about privacy when such data were going to be shared anonymously with researchers and the most concern when these data were going to be sold anonymously to third parties; the actual collection of the data by the app owners raised an intermediate level of concern. Finally, reviewing the same 12 types of data, we examined which types of data participants would be willing to share within future research studies—participants were most willing to share information they disclose publicly within the app (such as profile information on characteristics like age and height) and least willing to share information that could be collected by the app automatically (such as GPS location or device usage information).

We found overall moderate levels of trust within online research studies, with little difference based on the type of organization conducting the research. In this study, approximately two-thirds of GBMSM trusted or highly trusted online and mobile research compared with one-quarter who trusted GBMSM-specific networking apps and approximately 18% who trusted networking apps used by the general public. Although not asked in exactly the same way, these findings suggested lower levels of trust in this sample than those in a previous Pew poll [37], which could be due to the population or due to more general shifts that occurred in the year that passed between the 2 studies. Although we asked about trust in online research across three

different types of organizations (ie, LGBT community center, a university, and a government agency), participants in this study did not appear to differentiate between online research done by these three different groups and reported similar levels of trust for each. Particularly for researchers to understand, participants expressed greater trust in research than in many of the online and mobile technology companies and services they use on a regular—if not daily—basis. With the proliferation of public-private partnerships and collaborations between research organizations and these service providers, it is critical to consider how this might affect trust within both sources. Efforts to maintain trust by promoting transparency in research practices within such partnerships might prove critical. For example, getting informed consent before having data shared anonymously could be the best practice, even when such permission has already been granted within the terms of service for the app or site and research activities might qualify for a waiver of informed consent based on the federal criteria for human subjects review exemptions if data are transmitted anonymously.

In this study, participants expressed concern about several data collection, selling, and sharing activities. These findings are consistent with a study on the *privacy paradox* [43], which suggests that individuals' concerns about privacy are discrepant with their own privacy practices (eg, privacy settings). In this case, the paradox results from participants expressing concerns about the very types of data collection, selling, and sharing that they have likely already agreed to within the terms of service and privacy policies of the very apps and sites they regularly use and from which they were enrolled. Also, somewhat unexpectedly, more participants expressed concern about the actual collection of these various types of data by the app owners than they did about the anonymous sharing of the same data with researchers (an act that would be impossible without the apps first collecting these data). One potential explanation for this set of findings could be that the data remain connected to participants' identities for the app owners, whereas they were specifically referenced as anonymous when sharing with researchers. Another possibility is that this higher willingness to share data with researchers than have it collected in the first place by the app owners is due to the higher levels of trust in research that were observed within the analyses for the first aim of this study. Nevertheless, further investigation is warranted to explore the potential mediating roles of anonymity and trust on these differential privacy concerns. For example, privacy concerns might be lower for anonymous data activities than for identified ones—people may have similar or even higher levels of privacy concerns about sharing with researchers as they do about the app owners collecting the data if the sharing is not anonymous. Relatedly, people who trust different sources more might also express fewer privacy concerns, and so differences observed may be due to greater trust in research than the technology companies themselves.

Not surprisingly, similar types of data that participants expressed privacy concerns about were those that they were least willing to share with researchers. This might have implications for policies around broad consent for data sharing, whereby participants might need to be given the choice to opt in or out of specific types of data collection and sharing activities rather than simply consenting to share or not share all data. Specifically, these findings suggest that if individuals are given a choice of sharing all data or none, many might select to not share, resulting in low enrollment and high rates of missing data thus biasing the sample and study results. Alternatively, providing options about what to share might, at the very least, allow a more representative sample on some of the types of data (eg, sociodemographics) and could allow for a better estimate of how biased the results are for the types of data not shared. However, this study did not examine the impact of compensation, and further research is needed to examine how compensation might alter participants' willingness to engage in data sharing; understanding the impact of compensation on data sharing—particularly types of data that participants are otherwise generally unwilling to share—may inform ethical considerations.

Finally, data for this study were collected prior to the recent concerns about data-related and privacy issues on both social networking sites and sexual networking apps [44,45], and replication of the findings in the wake of ongoing privacy-related events is warranted. Future research can and should attempt to understand the magnitude and longevity of the impact of these events on constructs such as trust, privacy concerns, and willingness to share data. In the wake of such events, many technology companies seek to update their privacy policies but may do so with little information on what types of protections are most important to their users—researchers studying privacy in online and mobile technologies both within and outside of research are well-suited to understand and subsequently advise on exactly these types of issues.

Study Strengths and Limitations

In this study, we considered the use of technology and limited interaction procedures as strength as it facilitated large-scale data collection of individuals with substantially fewer resources than would be possible in a standard research study. However, it also necessitated conducting a very brief survey with a limited number of measures. We used a targeted advertisement with a random chance for incentives along with rigorously implementing standards for confirming the veracity and uniqueness of participants to reduce the likelihood of false and duplicate participants [41]. However, in online studies such as this, some degree of duplication or invalid response is likely. We reviewed the terms of service and other policies for several existing social media and sexual networking sites and apps while developing our measures to contextualize them appropriately. Nonetheless, the study constraints limited our ability to ask about the extent to which participants have ever read these policies, and the extent to which they realize that many of the data types and collection activities assessed were those that they have generally agreed to for apps and sites they regularly use remains unclear. We adjusted for rather than focusing on the role of sociodemographic and behavioral factors and future research is warranted to explore how trust, privacy, and willingness to share data might differ according to factors such as HIV status and race/ethnicity. Finally, conclusions regarding trust in this study and concerns about privacy are slightly limited as this is, by definition, a sample that agreed to

participate in an online research study. However, this is also a sample using the apps and sites from which they were enrolled, which were still trusted less than research, suggesting this finding regarding the relative trust could be reliable more generally even if the actual levels of trust are skewed higher by the nature of the sample.

Conclusions

This study suggests a relatively favorable view of online and mobile research—this large sample of GBMSM across the United States expressed a moderate level of trust in online research and few data-related privacy concerns. Moreover, the sample was nearly evenly split based on their willingness to have several types of app-based data shared with researchers, suggesting the analysis of such data might be potential avenues for future collaborations between researchers and technology companies. The findings highlighted the role of the privacy paradox, as participants expressed concerns about numerous data-related activities that they have likely permitted upon agreeing to use the apps and websites from which they were enrolled. Thus, researchers and ethical boards should consider these moderate levels of trust, privacy concerns, and willingness to share data when evaluating the risks and benefits of such partnerships. Meanwhile, other perspectives, such as legal and technical insights, should also be considered. When researchers can affect decision making, apps used for research purposes should be designed to decrease the extent to which participants must agree to data collection activities that concern them. For example, allowing participants to opt in or out of different aspects and providing multimedia (ie, “gist”) rather than text-based (ie, “verbatim”) explanations of the terms might reduce the privacy paradox in online and mobile research. For any secondary collection of data from apps, researchers should provide potential participants control over the types of data shared to the greatest extent possible, given the varying levels of concerns across different types of data that apps might have access to. Further research in this area is critical, particularly in the light of ongoing public awareness of and debate about technology and privacy [44,45].

Acknowledgments

HJR was supported in part by a career development award from the National Institute on Drug Abuse (K01-DA039060; PI: HJR). Data collection for this paper was supported in part by the Fordham HIV Prevention Research Ethics Training Institute (RETI) via a training grant sponsored by the National Institute on Drug Abuse (R25-DA031608, PI: Celia B Fisher). The authors also acknowledge the generous funding provided by the offices of the President, the Provost, and the Dean of Arts & Sciences of Hunter College, CUNY; additional support was also provided by Hunter College’s Center for HIV Educational Studies & Training (CHEST). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Fordham HIV Prevention Research Ethics Training Institute, or Hunter College, CUNY.

The authors would like to acknowledge the mentorship and feedback provided by the Fordham HIV Prevention Research Ethics Training Institute, particularly that of Dr Celia B Fisher and Dr Brenda Curtis. The authors also acknowledge the contributions of Dr Jeffrey Parsons and the CHEST Research Team, in particular those who played important roles in the implementation of the project: Ruben Jimenez, Chloe Mirzayi, and Scott Jones.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Measures of trust, privacy concerns, and data sharing.

[PDF File (Adobe PDF File), 68KB - jmir_v20i7e233_app1.pdf]

References

1. Grov C, Breslow AS, Newcomb ME, Rosenberger JG, Bauermeister JA. Gay and bisexual men's use of the Internet: research from the 1990s through 2013. J Sex Res 2014;51(4):390-409 [FREE Full text] [doi: 10.1080/00224499.2013.871626] [Medline: 24754360]
2. Liau A, Millett G, Marks G. Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. Sex Transm Dis 2006 Sep;33(9):576-584. [doi: 10.1097/01.olq.0000204710.35332.c5] [Medline: 16540884]
3. Mustanski BS. Getting wired: Exploiting the internet for the collection of valid sexuality data. Journal of Sex Research 2001 Nov;38(4):292-301. [doi: 10.1080/00224490109552100]
4. Rendina HJ, Jimenez RH, Grov C, Ventuneac A, Parsons JT. Patterns of lifetime and recent HIV testing among men who have sex with men in New York City who use Grindr. AIDS Behav 2014 Jan;18(1):41-49 [FREE Full text] [doi: 10.1007/s10461-013-0573-2] [Medline: 23925515]

5. Hirshfield S, Grov C, Parsons JT, Anderson I, Chiasson MA. Social media use and HIV transmission risk behavior among ethnically diverse HIV-positive gay men: results of an online study in three U.S. states. Arch Sex Behav 2015 Oct;44(7):1969-1978. [doi: 10.1007/s10508-015-0513-5] [Medline: 26179596]
6. Chiasson MA, Parsons JT, Tesoriero JM, Carballo-Dieguez A, Hirshfield S, Remien RH. HIV behavioral research online. J Urban Health 2006 Jan;83(1):73-85 [FREE Full text] [doi: 10.1007/s11524-005-9008-3] [Medline: 16736356]
7. Parsons JT, Vial AC, Starks TJ, Golub SA. Recruiting drug using men who have sex with men in behavioral intervention trials: a comparison of internet and field-based strategies. AIDS Behav 2013 Feb;17(2):688-699 [FREE Full text] [doi: 10.1007/s10461-012-0231-0] [Medline: 22684657]
8. Hernandez-Romieu AC, Sullivan PS, Sanchez TH, Kelley CF, Peterson JL, Del RC, et al. The comparability of men who have sex with men recruited from venue-time-space sampling and facebook: a cohort study. JMIR Res Protoc 2014;3(3):e37 [FREE Full text] [doi: 10.2196/resprot.3342] [Medline: 25048694]
9. Goedel WC, Duncan DT. Geosocial-Networking App Usage Patterns of Gay, Bisexual, and Other Men Who Have Sex With Men: Survey Among Users of Grindr, A Mobile Dating App. JMIR Public Health Surveill 2015;1(1):e4 [FREE Full text] [doi: 10.2196/publichealth.4353] [Medline: 27227127]
10. Holloway IW, Pulsipher CA, Gibbs J, Barman-Adhikari A, Rice E. Network Influences on the Sexual Risk Behaviors of Gay, Bisexual and Other Men Who Have Sex with Men Using Geosocial Networking Applications. AIDS Behav 2015 Jun;19 Suppl 2:112-122 [FREE Full text] [doi: 10.1007/s10461-014-0989-3] [Medline: 25572832]
11. Zou H, Fan S. Characteristics of Men Who Have Sex With Men Who Use Smartphone Geosocial Networking Applications and Implications for HIV Interventions: A Systematic Review and Meta-Analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
12. Rosengren AL, Huang E, Daniels J, Young SD, Marlin RW, Klausner JD. Feasibility of using Grindr to distribute HIV self-test kits to men who have sex with men in Los Angeles, California. Sex Health 2016 May 23;13:389-392. [doi: 10.1071/SH15236] [Medline: 27209076]
13. Rosenberger JG, Reece M, Novak DS, Mayer KH. The Internet as a valuable tool for promoting a new framework for sexual health among gay men and other men who have sex with men. AIDS Behav 2011 Apr;15 Suppl 1:S88-S90. [doi: 10.1007/s10461-011-9897-y] [Medline: 21331800]
14. Newcomb ME, Mongrella MC, Weis B, McMillen SJ, Mustanski B. Partner Disclosure of PrEP Use and Undetectable Viral Load on Geosocial Networking Apps: Frequency of Disclosure and Decisions About Condomless Sex. J Acquir Immune Defic Syndr 2016 Feb 01;71(2):200-206 [FREE Full text] [doi: 10.1097/QAI.0000000000000819] [Medline: 26761520]
15. Sexual and Gender Minority Research Office. National Institutes of Health. 2016. Minority research portfolio analysis URL: https://dpcpsi.nih.gov/sites/default/files/SGMRO_2016_Portfolio_Analysis_Final.pdf [accessed 2018-04-05] [WebCite Cache ID 6yRTm4EIA]
16. Pérez-Stable EJ. National Institute of Minority Health and Health Disparities. 2016. Sexual and gender minorities formally designated as a health disparity population for research purposes URL: https://www.nimhd.nih.gov/about/directors-corner/message.html [accessed 2018-04-05] [WebCite Cache ID 6yRU2BGLP]
17. Buchanan EA, Hvizdak EE. Online survey tools: ethical and methodological concerns of human research ethics committees. J Empir Res Hum Res Ethics 2009 Jun;4(2):37-48. [doi: 10.1525/jer.2009.4.2.37] [Medline: 19480590]
18. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and Privacy Implications of Using the Internet and Social Media to Recruit Participants for Health Research: A Privacy-by-Design Framework for Online Recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
19. Buchanan EA. Internet research ethics: Past, present, and future. In: Consalvo M, Ess C, editors. The handbook of Internet studies. West Sussex, UK: Wiley-Blackwell Publishing; 2011:83-108.
20. Curtis BL. Social networking and online recruiting for HIV research: ethical challenges. J Empir Res Hum Res Ethics 2014 Feb;9(1):58-70 [FREE Full text] [doi: 10.1525/jer.2014.9.1.58] [Medline: 24572084]
21. Allison S, Bauermeister JA, Bull S, Lightfoot M, Mustanski B, Shegog R, et al. The intersection of youth, technology, and new media with sexual health: moving the research agenda forward. J Adolesc Health 2012 Sep;51(3):207-212 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.06.012] [Medline: 22921129]
22. Ybarra ML, Prescott TL, Phillips GL, Parsons JT, Bull SS, Mustanski B. Ethical Considerations in Recruiting Online and Implementing a Text Messaging-Based HIV Prevention Program With Gay, Bisexual, and Queer Adolescent Males. J Adolesc Health 2016 Jul;59(1):44-49 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.03.020] [Medline: 27185621]
23. Arora S, Yttri J, Nilse W. Privacy and Security in Mobile Health (mHealth) Research. Alcohol Res 2014;36(1):143-151 [FREE Full text] [Medline: 26259009]
24. Labrique AB, Kirk GD, Westergaard RP, Merritt MW. Ethical issues in mHealth research involving persons living with HIV/AIDS and substance abuse. AIDS Res Treat 2013;2013:189645 [FREE Full text] [doi: 10.1155/2013/189645] [Medline: 24171110]
25. Capon H, Hall W, Fry C, Carter A. Realising the technological promise of smartphones in addiction research and treatment: An ethical review. Int J Drug Policy 2016 Oct;36:47-57. [doi: 10.1016/j.drugpo.2016.05.013] [Medline: 27455467]

26. S. Bhuyan S, Kim H, Isehunwa OO, Kumar N, Bhatt J, Wyant DK, et al. Privacy and security issues in mobile health: Current research and future directions. Health Policy and Technology 2017 Jun;6(2):188-191. [doi: 10.1016/j.hlpt.2017.01.004]
27. Martínez-Pérez B, de LTI, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
28. Vodicka E, Mejilla R, Leveille SG, Ralston JD, Darer JD, Delbanco T, et al. Online access to doctors' notes: patient concerns about privacy. J Med Internet Res 2013;15(9):e208 [FREE Full text] [doi: 10.2196/jmir.2670] [Medline: 24072335]
29. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. JMIR Mhealth Uhealth 2015;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
30. Allyse M, Karkazis K, Lee SS, Tobin SL, Greely HT, Cho MK, et al. Informational risk, institutional review, and autonomy in the proposed changes to the common rule. IRB 2012;34(3):17-19. [Medline: 22830179]
31. Park YJ, Mo Jang S. Understanding privacy knowledge and skill in mobile communication. Computers in Human Behavior 2014 Sep;38:296-303. [doi: 10.1016/j.chb.2014.05.041]
32. Fife E, Orjuela J. The Privacy Calculus: Mobile Apps and User Perceptions of Privacy and Security. International Journal of Engineering Business Management 2012 Jan;4(11):1-10. [doi: 10.5772/51645]
33. Keith MJ, Thompson SC, Hale J, Lowry PB, Greer C. Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior. International Journal of Human-Computer Studies 2013 Dec;71(12):1163-1173. [doi: 10.1016/j.ijhcs.2013.08.016]
34. Kehr F, Kowatsch T, Wentzel D, Fleisch E. Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. Info Systems J 2015 Mar 18;25(6):607-635. [doi: 10.1111/isj.12062]
35. Keith MJ, Maynes C, Lowry PB, Babb J. Privacy fatigue: The effect of privacy control complexity on consumer electronic information disclosure. 2014 Dec 14 Presented at: International Conference on Information Systems; 2014/12/14; Aukland, NZ URL: https://ssrn.com/abstract=2529606
36. Rainie L, Duggan M. Pew Research Center. 2016 Jan 14. Privacy and information sharing URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2016/01/PI_2016.01.14_Privacy-and-Info-Sharing_FINAL.pdf [WebCite Cache ID 6ySadFPz8]
37. Olmstead K, Smith A. Pew Research Center. 2017 Jan 26. Americans and cybersecurity URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2017/01/26102016/Americans-and-Cyber-Security-final.pdf [WebCite Cache ID 6ySakgLaf]
38. Mathieu E, Barratt A, Carter SM, Jamtvedt G. Internet trials: participant experiences and perspectives. BMC Med Res Methodol 2012;12:162 [FREE Full text] [doi: 10.1186/1471-2288-12-162] [Medline: 23092116]
39. Atienza AA, Zarcadoolas C, Vaughon W, Hughes P, Patel V, Chou WS, et al. Consumer Attitudes and Perceptions on mHealth Privacy and Security: Findings From a Mixed-Methods Study. J Health Commun 2015;20(6):673-679. [doi: 10.1080/10810730.2015.1018560] [Medline: 25868685]
40. Cheung C, Bietz MJ, Patrick K, Bloss CS. Privacy Attitudes among Early Adopters of Emerging Health Technologies. PLoS One 2016;11(11):e0166389 [FREE Full text] [doi: 10.1371/journal.pone.0166389] [Medline: 27832194]
41. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to “fraudsters” in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
42. Kirk R. Practical Significance: A Concept Whose Time Has Come. Educational and Psychological Measurement 2016 Jul 02;56(5):746-759. [doi: 10.1177/0013164496056005002] [Medline: 22973420]
43. Bavelloni A, Piazzi M, Raffini M, Faenza I, Blalock WL. Prohibitin 2: At a communications crossroads. IUBMB Life 2015 Apr;67(4):239-254 [FREE Full text] [doi: 10.1002/iub.1366] [Medline: 25904163]
44. Kang C, Frenkel S. Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users. NY, USA: The New York Times; 2018 Apr 4. URL: https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html [accessed 2018-06-08] [WebCite Cache ID 701ExG2Sh]
45. Singer N. Grindr Sets Off Privacy Firestorm After Sharing Users’ H.I.V.-Status Data. NY, USA: The New York Times; 2018 Apr 3. URL: https://www.nytimes.com/2018/04/03/technology/grindr-sets-off-privacy-firestorm-after-sharing-users-hiv-status-data.html [accessed 2018-06-08] [WebCite Cache ID 701FTSCyl]

Abbreviations

GBMSM: gay, bisexual, and other men who have sex with men
GPS: global positioning system
LGBT: lesbian, gay, bisexual, and transgender
RMANOVA: repeated measures analysis of variance

Edited by G Eysenbach; submitted 28.09.17; peer-reviewed by J Delgado-Ron, J Mitchell, S Hirshfield; comments to author 25.01.18; revised version received 05.04.18; accepted 12.05.18; published 04.07.18.

Please cite as:
Rendina HJ, Mustanski B
Privacy, Trust, and Data Sharing in Web-Based and Mobile Research: Participant Perspectives in a Large Nationwide Sample of Men Who Have Sex With Men in the United States
J Med Internet Res 2018;20(7):e233
URL: http://www.jmir.org/2018/7/e233/
doi:10.2196/jmir.9019
PMID:29973332

Tutorial

Creating Low-Cost 360-Degree Virtual Reality Videos for Hospitals: A Technical Paper on the Dos and Don'ts

Benjamin O'Sullivan[1,2], MBCHB; Fahad Alam[2,3], MD, MSc, MEd; Clyde Matava[1,2], MBChB, MMED

[1]Department of Anesthesia and Pain Medicine, Hospital for Sick Children, Toronto, Toronto, ON, Canada

[2]Department of Anesthesia, University of Toronto, Toronto, ON, Canada

[3]Department of Anesthesia, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

Corresponding Author:
Clyde Matava, MBChB, MMED
Department of Anesthesia and Pain Medicine
Hospital for Sick Children, Toronto
555 University Ave
Toronto, ON,
Canada
Phone: 1 8137445
Email: clyde.matava@sickkids.ca

Abstract

This article will provide a framework for producing immersive 360-degree videos for pediatric and adult patients in hospitals. This information may be useful to hospitals across the globe who may wish to produce similar videos for their patients. Advancements in immersive 360-degree technologies have allowed us to produce our own "virtual experience" where our children can prepare for anesthesia by "experiencing" all the sights and sounds of receiving and recovering from an anesthetic. We have shown that health care professionals, children, and their parents find this form of preparation valid, acceptable and fun. Perhaps more importantly, children and parents have self-reported that undertaking our virtual experience has led to a reduction in their anxiety when they go to the operating room. We provide definitions, and technical aspects to assist other health care professionals in the development of low-cost 360-degree videos.

(J Med Internet Res 2018;20(7):e239) doi:10.2196/jmir.9596

KEYWORDS

360-degree video; VR; virtual reality; video production; anesthetic preparation; preoperative anxiety; preoperative preparation

Introduction

This article aims to provide both methods and practical advice for the production of immersive 360-degree videos for children in hospitals. It is targeted at those with previous experience in the production of standard videos. Preoperative preparation of children is a well-researched method of reducing perioperative anxiety and the consequences of this anxiety [1]. The advancements in immersive 360-degree technologies have allowed for producing a "virtual experience" where children can prepare for anesthesia by actually "experiencing" all the sights and sounds of receiving and recovering from an anesthetic.

There are many standard videos available online aimed at the preparation of children for their hospital procedures. It would also be possible to produce these videos in panoramic or 180-degree modes, which would be technically much easier to produce. However, children and their parents prefer using a 360-degree video. The benefits of 360-degree video over these other methods are threefold [2,3]:

1. It provides a full visual account of what the child could see, there is less chance of surprise at the time of their anesthetic.
2. The increased autonomy for the child during the process of preparation may itself lead to reduced levels of anxiety on the lead up to their anesthetic.
3. The use of virtual reality (VR) headsets or 360 video viewers, by the very nature of being a toy, may reduce anxiety by the child associating the anesthetic with something fun.

Furthermore, it has been demonstrated that health care professionals, children, and their parents find this form of preparation valid, acceptable, and fun [2,3]. Perhaps more importantly, children, and parents have self-reported that undertaking our virtual experience has led to a reduction in their anxiety when they go to the operating room (OR) [2,3]. We provide definitions and technical advice to assist other health

care professionals in the development of low-cost 360-degree videos.

The production of 360-degree videos is significantly more challenging than the production of standard videos. There are many production steps which should be planned: (1) production of a script and recruitment of actors, (2) filming individual 360-degree scenes with an appropriate camera, (3) footage from each camera needs to be added together to produce the completed 360-degree video during a process called stitching, (4) the 360-degree footage needs to be edited into the completed film with appropriate software, and (5) the video file needs to be loaded onto a device supporting 360-degree video with a 360-degree viewer.

Production of the Script and Recruitment of Actors

Content and Pace of Video

As with planning the production of any video, it is essential to carefully consider what information the child will gain from the experience. It is important to optimize the length and pace of the video to ensure that it contains all the necessary information and does not result in symptoms such as motion sickness, dizziness, and headaches. An unnecessarily long video may increase the incidence of such symptoms or boredom, and the pace of the video should mimic real anesthetic experience as much as possible.

A key element in this is to perform a needs assessment to gain perspective from all stakeholders. This includes surveys, interviews, and a review of pre-existing standard videos created. Decision points for 360-degree videos will include deciding which elements of standard videos will work in that format.

Textbox 1 shows the basic structure of the video which lasts approximately 6 minutes. This serves as an example of using a needs assessment to inform the critical steps of the education 360 experience created for the viewer. The video was aiming to alleviate the anxiety of the preoperative experience. Textbox 1 has the 6 important phases that educate the viewer about this process. From a survey of 300 people (ie, 100 health care professionals, 100 parents, and 100 children), 291/300 (97%) rated the amount of information as optimal, and 288/300 (96%) rated the pace as optimal. Of particular note, none of the children desired any additional information nor would have liked any information to be deleted [2,3,4].

Actors

The virtual experience appears more realistic if it flows naturally through the individual scenes rather than shorter sequences being pasted together. For this reason, it is much easier to use professional actors or staff that are acting in their usual professional roles. Even the most experienced staff member may need several takes to get the scene completed perfectly when the camera is rolling.

Textbox 1. The structure of the virtual experience.

1. The preoperative area: Introduction and orientation from the nurse.
2. Walking down the corridor: The nurse explains where they are going and what they will see.
3. Walking into the OR: Meeting everyone in the OR and being asked to get onto the bed.
4. Lying on the operating bed: The nurse explains what she is doing while attaching all the routine monitors.
5. The anesthetist explains what they are doing while delivering a gas induction or inserting an intravenous (IV) tube and delivering an IV induction. The screen fades to black depicting falling asleep.
6. A brief period of darkness indicating being asleep.
7. In postoperative care unit: The nurse explains that procedure has finished, points out remaining pieces of monitoring, IV access, and reassures the child that the parents will be arriving soon.

How to Film a 360-Degree Video

Choosing a 360-Degree Camera

There exists an increasing number of 360-degree cameras available to the consumer. They all have different technical specifications, associated software, and workflows for stitching images. These vary in cost as well as using rigged or non-rigged cameras. A framework and rationale for the cameras selected is provided.

Rigged Cameras or Dual Lens Cameras

Two different set ups of cameras for shooting 360-degree videos were tested: (1) GoPro rig and (2) Ricoh Theta S. The first comprises a frame with 6 cameras fixed to it pointing along all 6 axes. The second was a self-contained system with two 180-degree lens cameras located back to back. Although the GoPro rigged system was superior concerning video quality, 3 significant disadvantages of this system over the self-contained 2 camera system were found. First, it was significantly bulkier and heavier which impeded setting up the camera in a number of the scenes. Second, there is footage from 6 cameras that need stitching together as opposed to 2 (dual lens camera), greatly increasing the postproduction work-load due to the increased amount of space between cameras and the amount of footage not captured, this set-up became inferior when filming objects at close range, such as face masks (Figure 1).

Figure 1. a) Optimal setup position of 360 camera; b) Optimal setup position of 360 camera with sheet covering the face; c) View through 360-degree viewer.

Camera Definition

Resolution of video footage can be considered as the amount of information the camera collects in each frame. It is essential to understand that to produce a clear picture in immersive videos, a higher definition is required. This is for two main reasons. First, each frame of the video is not being watched in its entirety on a flat screen in front of the user. The information in each frame is, in fact, being stretched around a central point into a full 360-degree environment leading to a significant loss of resolution (Multimedia Appendix 1 and Multimedia Appendix 2). Second, the video is viewed at a distance of about 2 inches from the user's eyes, within the 360-video viewer, and therefore any reduction in definition becomes magnified.

The first video was filmed using the Ricoh Theta S which claims a high-resolution of 1920 by 1080. Although the auto-stitch feature, which works well, the end product when viewed in the 360-degree viewer was of standard definition. 3/100the video complained of a blurry picture, sore eyes, and a headache. This was attributed to the low resolution of the video [3].

For the second video, the Samsung Gear 360 was used which has a higher video resolution of 3840 by 1920. Although this device does not come with an auto-stitch feature, it can produce a much clearer higher definition immersive experience. Although not formally assessed, the higher resolution appears to cause less incidence of sore eyes and headaches.

How to Obtain a 360-Degree Footage

When producing non 360-degree films, the footage can be viewed as it is filmed. With 360-degree filming, everything in the line of sight of the camera is recorded, and it is not until after the stitching and into the editing process that the scene can be viewed entirely. It is a worthwhile exercise to have dry runs of each scene with someone in the position of the camera looking around the entire room to see exactly what the camera will see during the filming. This is especially important when filming in a hospital environment as it is easy to inadvertently film sensitive material such as a patient in the background or a piece of patient identifiable data on a computer screen.

The optimal way to film the experience was to use the camera as the head of the "patient" and administer an anesthetic to the camera. This ensured that the scene felt real and also enabled the user to feel like they are being spoken to and interacted with personally during their virtual experience.

To film the torso of the child and depict the application of various physiological monitors and lines, it is possible to either use a mannequin or an actor. The method we describe here uses a mannequin, but the same principles may be applied to a live model.

The camera needs to be positioned on the mannequin so that the child sees the "body" as they would their own. After testing different set-up options, the optimal position to set up the camera is as depicted in Figure 1a and b. This enables the child to see everything that happens during the induction of anesthesia and the application of monitoring equipment as shown in Figure 1c.

One aspect of preanesthetic preparation for children which the virtual experience lends itself particularly well to is the gas induction. Holding the top of the mask 3-4 inches away from the camera and 1 inch above the center of the camera allows the child to experience how the mask will appear during a gas induction, as shown in Figure 2a. This allows the child to get the claustrophobic impression of the mask but also allows them

still to see everything that is going on in the room. This is shown in Figure 2b. Children who have used the virtual experience have found this aspect of the preparation particularly useful.

One inherent technical problem was discovered when filming in this manner. Using the camera as the head of a body means that when the user looks down during the video, they will either see a face or a "headless" body depending on whether or not the face is covered. After multiple tests, it was discovered the most aesthetically pleasing method was to do 2 things. First, during the filming process, cover the face of the mannequin in a sheet the same color as the bed, as shown in Figure 1b. Second, during the editing process, place a logo at the bottom of the video to hide the space where the head would be located (Figure 3).

The scene depicting the walk to the OR posed another technical difficulty. In the first attempt, the nurse escorting the patient to the OR pushed an IV stand with the camera attached to the top. This initially appeared to work well; however, it did reveal one of the side effects caused by this technology. During the virtual experience, the user remains in a sitting position and is placed in a "virtual environment" where they perceive motion, which can lead to motion sickness. The first video caused 20/57 (35%) of users to feel dizzy and 10/57 (17.5%) to feel nauseous. Subjects attributed their side effects predominantly to this scene. The scene was refilmed using an IV stand with added weights at the bottom, to increase the stability of the pole, and wheeled it down the corridor much more slowly at a speed of about 0.25meters/second. This enabled the scene to be produced where the user still gets the impression of walking down the corridor but has a much lower incidence of motion sickness. The downside to this is that it was only possible to film a small portion of the walk to the OR as it would have taken too long to walk at this speed.

Obtaining Audio for a 360-Degree Video

There are multiple ways of recording the audio for the video. The simplest option is to use the built-in microphone in the camera. This, however, does not provide the best audio quality, especially in an environment with lots of ambient noise. Voices from different locations within the environment will also be recorded at different volumes.

The most reliable way to record the actors' voices is to fit each with microphones and add this to the video at the editing stage. Regarding background sounds, which are important for the experience, such as monitoring sounds or trays being opened, it is best to film the experience in a quiet room. Separately record these sounds and then add in during the editing. It is also possible to record actors' voices at this stage.

Figure 2. a) Optimum position of the mask during a gas induction; b) View of this setup through the 360-degree viewer.

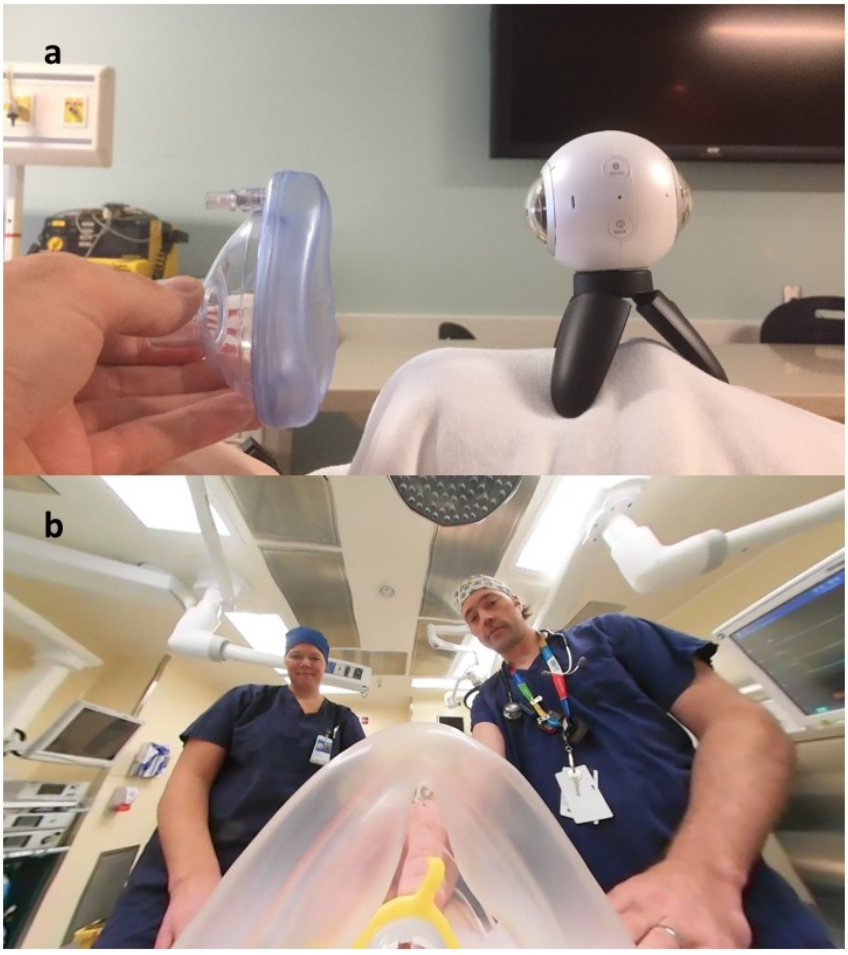


Figure 3. View through 360-degree viewer when looking down at our hospital's logo.

Stitching the Video Footage

This is a very time consuming and technically challenging task. Each frame from each camera needs to be stitched together to produce a video file that contains a full 360-degrees worth of information. The Ricoh Theta S camera allowed for a bypass to this stage as it had an auto stitch function enabling the export of data directly from the camera to the editing software. The Samsung Gear 360 did not come with this function, and it was necessary to obtain help from an outside company to complete the stitching process. This is a very significant consideration when deciding on which camera to use, as it may cost significantly more to hire outside help to complete the stitching of the footage.

Editing and Combining Footage Into a Full Video

The editing techniques behind producing a completed video are beyond the scope of this article, and there are many editing software packages available. However, some of the fundamental concepts when editing a 360-degree video are explained here. The editing software used was Adobe Premiere.

Equirectangular Videos

After the stitching process has been completed, the video files are in equirectangular format. This means that the 360-degree spherical images have been flattened and distorted onto a 2-dimensional rectangle, much like a map of the world. Therefore, when viewing the clips during the editing process, they will appear distorted, and it is difficult to appreciate what the final product will look like (Figure 4).

Video Clip Settings

For the final product to work as required in the 360-degree video viewer, it is vital to maintain the correct relative dimensions of the video during the import process into the editing software and during the export process out of the editing software. The exact magnitude of the dimensions will vary depending on the resolution of the clips but will need to be twice as wide as they are high (ie, 3840 by 1920). It is also vital to ensure that there are no borders to the image as the edges of the video will be wrapped around and brought together when playing on the 360-degree video viewer.

Lighting

Even though additional lighting was used during the filming, the finished product was still quite dark. Increasing the exposure setting by approximately 40% and the saturations setting by approximately 80% on the editing software produced a video that is much brighter and warmer. It also gives a much clearer picture. The exact degree to which these settings should be adjusted will depend on the original footage.

Exporting Video From the Editing Software and Injecting Metadata

Table 1 shows the export settings that were used. The VR mode required will depend on the device used to view the final product. Most 360-degree editing software will automatically format the video file into a stereoscopic mode (ie, 2 images side by side) which is required to view the video file in the viewer. If software is used that does not do this, then it will be necessary to export as a stereoscopic video file.

Metadata

Metadata is the information that is embedded into the video file that allows the video to be viewed in 360-degree mode. Whether or not the raw footage contains this data depends on the camera used. It may be necessary to install this metadata during the export process. Using this method, it is not necessary to do this. However, it is critical to ensure that when the video is exported from the editor that the metadata remains within the video file by enabling the metadata settings.

Figure 4. How 360-degree footage appears when flattened onto a 2-Dimensional workspace.

Table 1. Video export settings.

Attribute	Setting
Format	H.264
Width	3840
Height	1920
Frame Rate	29.97
Aspect	Square Pixels
Bitrate	Maximum Bitrate
Virtual reality mode	Monoscopic
Metadata	Enabled

The 360-Degree Enabled Devices and Viewers

There are a large number of devices and viewers (eg, HTC Vive, Samsung Gear, Google Cardboard), that allow viewing 360-degree video media. The most reliable and cost-effective way of viewing the completed video is as follows. First, download the completed video to a mobile phone. All current mobile phones have the ability to run the necessary VR applications. Second, open the video in an appropriate VR application. The free and open source VR application called Childlife VR was used. Third, view the video through a 360 viewer. These are readily available devices for mobile phones. The Google Cardboard viewer was used in this study due to its availability and price.

For safety reasons, it must be ensured that all users of the video remain seated at all times during the experience. This still enables them to look all around and experience the full benefit of the video. Although most of the information happens in the "looking forward" position, children still choose to take advantage of the technology and look all around them throughout the video.

Discussion

This tutorial provides an account of the technical challenges encountered and techniques that were found to be effective when producing a video (ie, Childlife VR). There are many different ways of preparing children for a hospital procedure and not all will wish to participate in this particular way. There was 1/ 101 (1%) of the children approached to use the video chose not to try it as they had received over 8 previous anesthetics. They had their own coping mechanism and did not want this process interfered with. In addition, 6/100 (6%) children recruited preferred to use the standard methods of preparation in the future [1,2].

Three-dimensional cameras and augmented reality goggles are becoming more readily available and more reasonably priced. This may represent future alternative methods used in the preparation of children for general anesthesia and other hospital procedures.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of rigged set up (left) with dual camera (right). (A) shows the areas not captured by the cameras; (B) shows the areas that need "stitching" together

[PNG File, 94KB - jmir_v20i7e239_app1.png]

Multimedia Appendix 2

How footage is "stretched" around the user in order to be viewed in 360-degrees.

[PNG File, 48KB - jmir_v20i7e239_app2.png]

References

1. Capurso M, Ragni B. Psycho-educational preparation of children for anaesthesia: A review of intervention methods. Patient Educ Couns 2016 Feb;99(2):173-185. [doi: 10.1016/j.pec.2015.09.004] [Medline: 26603504]
2. O'Sullivan B, Brazel K, Alam F, Salman M, Caldeira M, Matava C. Virtual Reality for Educating and Reducing Preoperative Anxiety in Children-Phase 2: Design, Face Validity and Acceptability by Children and Their Parents. In: Anesthesia and Analgesia. Two Commerce Sq, 2001 Market St, Philadelphia, PA 19103 USA: Lippincott Williams & Wilkins; 2017:901-901.
3. O'Sullivan B, Brazel K, Caldeira M, Salman M, Alam F, Matava C. Virtual Reality for Educating and Reducing Preoperative Anxiety in Children-Phase 1: Design, Face Validity and Acceptability by Healthcare Professionals. Two Commerce Sq, 2001 Market St, Philadelphia, PA 19103 USA: Lippincott Williams & Wilkins; May 1, 2017:899-900.
4. Eijlers R, Legerstee JS, Dierckx B, Staals LM, Berghmans J, van DSMP, et al. Development of a Virtual Reality Exposure Tool as Psychological Preparation for Elective Pediatric Day Care Surgery: Methodological Approach for a Randomized Controlled Trial. JMIR Res Protoc 2017 Sep 11;6(9):e174 [FREE Full text] [doi: 10.2196/resprot.7617] [Medline: 28893727]

Abbreviations

IV: intravenous
OR: operating room
VR: virtual reality

Edited by G Eysenbach; submitted 20.12.17; peer-reviewed by T Guetterman, P Rhienmora, S Berrouiguet, S Mosadeghi, R Umoren; comments to author 23.01.18; revised version received 12.06.18; accepted 21.06.18; published 16.07.18.

Please cite as:
O'Sullivan B, Alam F, Matava C
Creating Low-Cost 360-Degree Virtual Reality Videos for Hospitals: A Technical Paper on the Dos and Don'ts
J Med Internet Res 2018;20(7):e239
URL: http://www.jmir.org/2018/7/e239/
doi:10.2196/jmir.9596
PMID:30012545

Original Paper

Electronic Health Literacy Across the Lifespan: Measurement Invariance Study

Samantha R Paige[1,2], MPH; M David Miller[3], PhD; Janice L Krieger[2,4,5], PhD; Michael Stellefson[6], PhD; JeeWon Cheong[1], PhD

[1]Department of Health Education and Behavior, University of Florida, Gainesville, FL, United States

[2]STEM Translational Communication Center, University of Florida, Gainesville, FL, United States

[3]School of Human Development and Organizational Studies in Education, University of Florida, Gainesville, FL, United States

[4]Department of Advertising, University of Florida, Gainesville, FL, United States

[5]Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States

[6]Department of Health Education and Promotion, East Carolina University, Greenville, NC, United States

Corresponding Author:
Samantha R Paige, MPH
STEM Translational Communication Center
University of Florida
PO Box 118400
Gainesville, FL, 32611
United States
Phone: 1 352 392 9961
Email: paigesr190@ufl.edu

Abstract

Background: Electronic health (eHealth) information is ingrained in the healthcare experience to engage patients across the lifespan. Both eHealth accessibility and optimization are influenced by lifespan development, as older adults experience greater challenges accessing and using eHealth tools as compared to their younger counterparts. The eHealth Literacy Scale (eHEALS) is the most popular measure used to assess patient confidence locating, understanding, evaluating, and acting upon online health information. Currently, however, the factor structure of the eHEALS across discrete age groups is not well understood, which limits its usefulness as a measure of eHealth literacy across the lifespan.

Objective: The purpose of this study was to examine the structure of eHEALS scores and the degree of measurement invariance among US adults representing the following generations: Millennials (18-35-year-olds), Generation X (36-51-year-olds), Baby Boomers (52-70-year-olds), and the Silent Generation (71-84-year-olds).

Methods: Millennials (N=281, mean 26.64 years, SD 5.14), Generation X (N=164, mean 42.97 years, SD 5.01), and Baby Boomers/Silent Generation (N=384, mean 62.80 years, SD 6.66) members completed the eHEALS. The 3-factor (root mean square error of approximation, RMSEA=.06, comparative fit index, CFI=.99, Tucker-Lewis index, TLI=.98) and 4-factor (RMSEA=.06, CFI=.99, TLI=.98) models showed the best global fit, as compared to the 1- and 2-factor models. However, the 4-factor model did not have statistically significant factor loadings on the 4th factor, which led to the acceptance of the 3-factor eHEALS model. The 3-factor model included eHealth Information Awareness, Search, and Engagement. Pattern invariance for this 3-factor structure was supported with acceptable model fit (RMSEA=.07, $\Delta\chi^2$=*P*>.05, ΔCFI=0). Compared to Millennials and members of Generation X, those in the Baby Boomer and Silent Generations reported less confidence in their awareness of eHealth resources (*P*<.001), information seeking skills (*P*=.003), and ability to evaluate and act on health information found on the Internet (*P*<.001).

Results: Young (18-48-year olds, N=411) and old (49-84-year olds, N=419) adults completed the survey. A 3-factor model had the best fit (RMSEA=.06, CFI=.99, TLI=.98), as compared to the 1-factor, 2-factor, and 4-factor models. These 3-factors included eHealth Information Awareness (2 items), Information Seeking (2 items), and Information and Evaluation (4 items). Pattern invariance was supported with the acceptable model fit (RMSEA=.06, $\Delta\chi^2$=*P*>.05, ΔCFI=0). Compared with younger adults, older adults had less confidence in eHealth resource awareness (*P*<.001), information seeking skills (*P*<.01), and ability to evaluate and act upon online health information (*P*<.001).

Conclusions: The eHEALS can be used to assess, monitor uniquely, and evaluate Internet users' awareness of eHealth resources, information seeking skills, and engagement abilities. Configural and pattern invariance was observed across all generation groups in the 3-factor eHEALS model. To meet gold the standards for factor interpretation (ie, 3 items or indicators per factor), future research is needed to create and assess additional eHEALS items. Future research is also necessary to identify and test items for a fourth factor, one that captures the social nature of eHealth.

(J Med Internet Res 2018;20(7):e10434) doi:10.2196/10434

KEYWORDS

eHealth literacy; eHealth; aging; measurement invariance

Introduction

Background

Telemedicine and electronic health (eHealth) transcends geographic, social, and political boundaries, making them essential tools to leverage health care delivery and surveillance [1,2]. The Internet has become deeply penetrated into society, with nearly 90% of adults in the United States having Internet access [3]. Millennials (18-35-year-olds), however, continue to adopt the Internet at a more rapid rate than members of Generation X (36-51-year-olds), Baby Boomers (52-70-year-olds), and the Silent Generation (71-115-year-olds) [4]. Although age-related disparities in Internet adoption have declined in recent years [5], the strategies to narrow this chasm and optimize the eHealth experience will require a closer look at the unique attributes of generations.

Generational differences in technological adoption can be broadly attributed to the point in one's life that technology was penetrated into society [6]. Members of Generation X created the same technology that has become central to Millennials' everyday lives. Rather than being familiar with and growing up with technology, Baby Boomers and members of the Silent Generation were introduced to technology after their social and cultural identities had been established. Widespread adoption of the Internet and the capabilities of technology have led Baby Boomers and members of the Silent Generation, who are traditionally considered late adopters of innovations like technology [4,6], to become excited and willing to adapt and learn about new technologies [7]. However, barriers related to unfamiliarity and uncertainty surrounding the use, value, and security of health information technologies persist among middle-to-older age adults [8-11], especially among those who are not avid health service users [12]. Evidence also shows that non-primary care physicians who are 55 years old and over are less likely to integrate electronic health record systems into their practice, as compared to their younger physician counterparts [13]. Consistent with theoretical underpinnings of the Diffusion of Innovation and the Technology Acceptance Model [7,14], technology tends to be adopted more quickly among younger age groups who find that it is both useful and easy to use.

Adoption of eHealth, however, does not ensure that the technology is used appropriately or that it is used to access high-quality and actionable health information [15]. eHealth literacy, driven by health and computer literacies, is defined as the capacity to locate, understand, evaluate, and act upon health information from technology [16]. People with a low degree of eHealth literacy are less likely to find the Internet as a useful health information tool, to trust the health information from diverse online sources and channels [17], and to actively seek out health information from the Internet [18]. Literacy in eHealth is a central skill set that influences not only health information seeking behaviors [19-21], but also the likelihood of engaging in proactive health-related outcomes and experiences [18,22]. Similar to generational values, researchers argue that social and cultural contextual frames influence eHealth literacy [23,24]. As such, understanding how generational age serves as a function of eHealth literacy and optimizing its measurement across these groups will be critical in the evolving technological era.

Empirical evidence over the past decade has shown that an inverse correlation exists between age and eHealth literacy [18,22,25,26]. Older adults generally have lower health literacy than their younger counterparts [27,28], yet this population is increasingly adopting the Internet with a high degree of confidence to access health information and supplement their health care [4,25]. Paige and colleagues [17] found that adults in the middle-to-older adult age groups, or Boomers and members of the Silent Generation, were more likely to have low eHealth literacy than their younger counterparts. Older age groups were also less likely to trust health information from social support forums but more likely to trust health information from Facebook. These age disparities have been attributed to older adults' unique health needs as compared to younger adults [29,30], including specialized health information related to chronic disease [31-33], the potential risk for social isolation [34,35], and physical and cognitive limitations that are due to the natural aging process [36]. The generational differences in information seeking behaviors in the non-health context have also been highlighted in the literature to show that Millennials and Baby Boomers consult different informational sources [37]. For these reasons, it would be naïve to assume that eHealth literacy is measured and conceptualized equivalently across generational age groups. To our knowledge, evidence to support that measurement invariance of eHealth literacy scores exists across generations does not exist.

Valid age-group comparisons of eHealth literacy and associated patient-reported outcomes cannot be established without evidence that eHealth literacy measures function equally, or invariant, among young and old adults [38,39]. Measurement invariance indicates that the latent construct captured by an instrument will function similarly across different groups. Multi-group comparisons that do not meet the assumption of measurement invariance are ambiguous and subject to bias [38,40] and may use misleading or false claims to advance

research and practice [41]. Without such evidence, it is unknown if the different relationship between eHealth literacy and age is due to real differences or systematic biases. As such, older adults may have a lesser degree of confidence to use eHealth. However, it is also possible that normal age-related cognitive declines [42] and low health literacy [28] readily reported among older adults contribute to depleted attention and working memory to recall accurate responses. As such, specific items may appear more salient in one age group over another. Establishing measurement invariance across eHealth literacy scales will have significant implications for fair and equitable testing standards. Also, it will alleviate bias in using these instruments to identify patients who are likely to benefit from online programs.

Since the conceptualization of eHealth literacy in 2006, several instruments have been developed to capture this construct in the evolving era of eHealth. The seminal instrument, the eHealth Literacy Scale (eHEALS), is a brief 8-item measure with theoretical underpinnings in self-efficacy, or the confidence in one's capability to engage in behavior to result in the desired outcome [43]. Alongside the emergence of online social environments, like social media, there have been criticisms that the instrument has a compromised degree of content validity [16], particularly regarding its insufficient ability to capture the multidimensional and dynamic features of eHealth. In response, formative research was conducted to derive constructs salient to eHealth literacy and its measurement inductively. The most significant contribution noted by these instruments is the ability to capture the dynamic feature of eHealth and pressing issues related to eHealth use (eg, privacy). These instruments included items that assess if an Internet user can talk to their offline health care provider about the health information found on the Internet [44], as well as their skills related to privacy protection and message self-creation with a keyboard [45]. Another instrument related to eHealth literacy assessed the Internet users preferred mode of interaction and online experiences, as well as their degree of computer anxiety and health information needs [46]. These new instruments tap into unique aspects of eHealth literacy, but they do not provide insight into the communication exchange processes that are missing from eHEALS. Instead, these instruments have been said to leave eHealth literacy literature static, as recent attempts to advance the concept and measurement have not built upon previous literature [47]. Given this information, it is possible that new operational definitions, concepts, and measures that do not build upon seminal work of eHealth literacy may lead researchers astray from the core operational behaviors (ie, locate, understand, evaluate, act upon). Although measures of eHealth literacy have been published, eHEALS remains the most widely used and refined instrument in the literature [48-50] .

Evidence for the internal structure and external validity of eHEALS as a unidimensional measure exists across diverse age groups. These populations include adolescents [51], college students [52], the general adult population [52], patients with chronic disease [53], older adults recruited to surveys conducted online [54], and baby boomers and older adults recruited through the telephone [55]. More recently, studies applying sophisticated psychometric modeling techniques have found that eHEALS is a multi-dimensional measure that captures operational behaviors consistent with the seminal operational definition of eHealth literacy [16]. The eHEALS has been identified as a 2-factor measure of eHealth literacy among Australian adults who are at-risk for cardiovascular disease [56]. The 2-factor model was replicated among general adult populations in Germany [57] and Israel [57,58]. These factors have been defined as measuring information seeking and information appraisal. Most recently, the eHEALS 3-factor structure has been reported among adults later in the lifespan. Sudbury-Riley and colleagues [49] report that eHEALS scores produce a 3-factor model of information awareness, seeking, and appraisal skills among baby boomers. Similarly, the 3-factor model of eHEALS has been confirmed among baby boomers and older adults [55], chronic disease patients [53], as well as middle age adults, with an average age of 53, in a magnetic resonance imaging and computed tomography medical imaging outpatient clinic [59]. The 3-factor structure of eHEALS scores, however, has not been reported or confirmed among younger age groups, like Millennials or members of Generation X. As such, the discrepancy in 1-, 2-, and 3-factors captured among middle-to-older age adults and the general population in international contexts brings into question whether or not eHEALS produces similar factor structures across the lifespan.

Sudbury-Riley and colleagues [49] found measurement invariance for the 3-factor structure among baby boomers in the US, United Kingdom, and New Zealand. As such, the multidimensional eHEALS structure does not vary among Baby Boomers across international borders, regardless of the various health care provisions and coordination that drive social and cultural frames. Age, however, is also a strong determinant that shapes and influences the social and cultural frame of a given population [60]. Baby Boomers are a single generation within the lifespan, and measuring their health-related technological skills is well justified. However, Baby Boomers are a single generation whose socio-cultural and political frame has a significant influence on health outcomes and health services uptake [61-63]. Evidence that research on generational differences in eHealth adoption sets a precedent to also consider the potential generational variability in measures that assess eHealth literacy.

Although a "gold standard" eHealth literacy instrument does not exist [47], the eHEALS remains the closest to reaching this status due to its brevity, popularity, and theoretical underpinnings in health behavior change theory. Researchers have recommended the refinement of the eHEALS, specifically to account for the social nature of eHealth [47,64,65]. Before embarking on this mission, there is an obligation to understand the multidimensional factor structure of the eHEALS across age groups and whether or not these factor structures are invariant. Without such evidence, it will be challenging to refine eHEALS as a reliable measure that produces scores with a high degree of validity evidence across the lifespan in the social era of eHealth. Therefore, the purpose of this study is to examine the structure of eHEALS scores and the degree of measurement invariance among three generations: Millennials, Generation X, and Baby Boomers and the Silent Generation.

Methods

Sample and Procedures

A sample of Qualtrics Panelists from the United States completed an online survey in May 2015, which was approved by the university Institutional Review Board (IRB). The sample was stratified by race (ie, Caucasian, Black/African American). Per Qualtrics Panels, the survey functioned as opt-in, meaning that panelists who met inclusion criteria were offered the opportunity to consent to participate. For this particular survey, the inclusion criteria included residing within the United States and being older than 18 years old. Some respondents (n=11) did not provide their age and were subsequently removed from the final sample (N=829). Upon removing respondents who did not provide an age value, there were no missing eHEALS data in this sample.

Measures

The following sociodemographic factors were measured across the sample [66]: (1) age (in years), (2) gender (Male, Female), (3) race (Black/African American, Caucasian), (4) ethnicity (Hispanic, non-Hispanic), (5) education level, (6) annual income, and (7) Internet use for Health. The eHealth Literacy was assessed with the eHEALS (Norman and Skinner, [51]), an 8-item, 5-point Likert-type rating scale (1=strongly disagree, 5=strongly agree).

Data Analysis

The age group of the sample was categories as: (1) Millennials (18-35-year-olds), (2) Generation X (36-51-year-olds), and (3) Baby Boomers (52-70-year-olds) and Silent Generation (71-115-year-olds) [67]. Baby Boomers and the Silent Generation were collapsed into a single group for this study because the sample only contained 45 members of the Silent Generation. Frequency and descriptive statistics were computed to describe the sample and eHealth literacy scores. A series of chi-squared analyses were conducted to determine if sociodemographic factors were significantly different by age group.

Dimensionality

Multi-group exploratory structural equation models (MG-ESEM) were conducted with Mplus v7.3 [68] to inform the number of factors underlying eHEALS items. MG-ESEM is not a confirmatory factor analysis approach. Instead, it is a structural equation modeling (SEM) approach that integrates principles of exploratory factor analysis. This statistical approach is justified by the limited, and inconsistent (eg, 1-factor, 2-factor, 3-factor), knowledge regarding measurement properties of eHEALS across generations. For example, there is no priory theory to support that a certain number of factors are salient across generations who complete the eHEALS. Moreover, there is limited theoretical support to suggest that specific items from eHEALS would belong to one factor over another. Considered a novel framework to examine the measurement and structural properties through a SEM lens [69], a similar statistical approach has been used in measurement studies to examine the properties of eHEALS among baby boomers and older adults [55].

The fit of 4 MG-ESEM factor models was evaluated across each generation. Each model was independently examined, beginning with 1-, 2-, 3-, and finally a 4-factor model structure. The following criteria were used to evaluate the global model fit of each model [70]: (1) statistically non-significant chi-square value, (2) Root mean square error of approximation (RMSEA) value less than .08, (3) comparative fit index (CFI) value greater than .95, (4) Tucker-Lewis index (TLI) value greater than .95, (5) standardized root mean square residual (SRMR) less than .08, and (6) smaller Akaike information criterion (AIC). Factor loadings of .30 or greater [71] were examined to form the base model for testing measurement invariance.

Measurement Invariance

Mplus v7.3 [68] was used to carry out 3 analytical invariance tests within Confirmatory Factor Analyses (CFA) framework to test if the instrument functions similarly across each generational age group [39]. There were 3 tests of measurement invariance conducted [72]. The first was *configural invariance*, in which all parameters from the factor model identified in the MG-ESEM are freely estimated across groups to confirm that the underlying factor structure is equivalent. Next, *pattern invariance* tests the equivalence of unstandardized factor loadings across groups, which is used to examine if items are related to the factors in the similar ways across groups. Finally, the *unique invariance* test examines the equivalence of item measurement error across groups. Chi-square difference tests were conducted for model comparisons to test each level of measurement invariance, and several fit indices such as RMSEA, SRMR, CFI, and TLI, were also examined to evaluate the fit of the final model. A change in chi-square statistic was compared to the critical value with the relevant for the change in degrees of freedom. If the chi-square difference test was significant, adding invariance constraints was considered worsening the model fit and indicating lack of invariance. As chi-square is sensitive to sample sizes [73], a CFI change less than .01 was considered as non-significant changes in model fit, supporting invariance [74].

Comparing Electronic Health Literacy Scale Scores by Age Group

The statistical software SPSS v24 [75] was used to examine the internal consistency, or Cronbach's alpha, of items comprising each eHEALS factor and compute the average of item scores. The reliability of each factor across age groups was determined by the omega coefficient, which is more appropriate for congeneric factor analysis models that do not function under tau-equivalence [76,77]. A one-way analysis of variance (ANOVA) and Tukey post-hoc analyses were conducted to identify the mean difference in eHealth literacy scores among each generation. Statistical significance was detected at $P<.05$.

Results

Sample Characteristics

As shown in Table 1, the mean ages of Millennials, Generation X, and Baby Boomers/Silent Generation Members were 26.64 (SD 5.14), 42.97 (SD 5.01), and 62.80 (SD 6.66), respectively. Respondents were mostly female (603/829, 72.74%), earning

at least US $35,000 each year (499/829, 60.41%), and living with at least some college experience (623/829, 75.15%%). There were no statistically significant differences in gender, income, or education across each generation group. Nearly half of the respondents were Black/African American (412/829, 49.70%) or Caucasian (417/829, 50.30%), and most were non-Hispanic or Latino (807/829, 97.2%). A greater number of Millennials used the Internet for health-related purposes, as compared with members of Generation X or Baby Boomers/Silent Generation groups (*P*=.009).

Dimensionality

The estimates of model fit for 1-4 factor models are presented in Table 2. Exceeding the acceptable level of RMSEA were the 1-factor (value=.14), 2-factor (value=.09) models. The 3-factor model (RMSEA=.06, 90% CI 0.04-0.08, CFI=.98, TLI=.98) and 4-factor model (RMSEA=.06, 90% CI 0.04-0.08, CFI=.99, TLI=.99) indicated good global model fit. Similarly, the AIC values for the 3-factor (value=12750) and 4-factor (value=12737.50) models were lower than the values for the 1-factor, 2-factor, and 4-factor models.

Table 1. Sociodemographics of millennials, generation X, and baby boomers/silent generation members.

Characteristic	Millennials (N=281)	Generation X (N=164)	Baby Boomers/Silent Generation (N=384)
Age in years, mean (SD)	26.64 (5.14)	42.97 (5.01)	62.80 (6.66)
Gender, n (%)			
Male	73 (25.9)	54 (32.9)	99 (25.7)
Female	207 (73.7)	110 (67.1)	286 (74.3)
Missing	1 (0.36)	0 (0.0)	0 (0.0)
Race[a], n (%)			
Black/African American	156 (55.7)	93 (56.7)	163 (42.3)
Caucasian	124 (44.3)	71 (43.3)	222 (57.7)
Ethnicity, n (%)			
Hispanic	8 (2.9)	2 (1.2)	6 (1.6)
Non-Hispanic	267 (95.4)	161 (98.2)	378 (98.2)
Missing	5 (1.8)	1 (0.6)	1 (0.3)
Education level, n (%)			
< High school	12 (4.3)	5 (3.0)	12 (3.1)
High school/GED	58 (20.7)	31 (18.9)	88 (22.9)
Some college	99 (35.4)	47 (28.7)	136 (35.3)
Bachelor’s degree	65 (23.2)	39 (23.8)	67 (17.4)
Master’s degree	26 (9.3)	28 (17.1)	59 (15.3)
Advanced graduate	18 (6.4)	13 (7.9)	23 (5.9)
Missing	2 (0.7)	1 (0.6)	0 (0.0)
Annual income (US $), n (%)			
≤$20K/year	60 (21.6)	28 (17.1)	65 (16.9)
$20K-$34,999K/year	62 (22.3)	33 (20.1)	79 (20.6)
$35K-$49,999K/year	50 (18)	27 (16.5)	61 (15.9)
$50K-$74,999K/year	57 (20.5)	25 (15.2)	88 (22.9)
≥$75K more/year	49 (17.6)	51 (31.1)	91 (23.7)
Internet use for health[b], n (%)			
Yes	278 (99.3)	157 (95.7)	366 (95.1)
No	2 (0.7)	7 (4.3)	19 (4.9)

[a]Black/African Americans and Caucasian respondents were less likely to be a member of Generation X than any other generation, $\chi^2(2, N=829)=15.62$, *P*<.001.

[b]More Millennials reported using the Internet for health, as compared to members of Generation X or Baby Boomers/Silent Generation, χ^2 (2, N=829)=9.35, *P*=.009.

Table 2. Global model fit estimates for multi-group exploratory structural equation models

Model	χ^2 *(df)*	*P* value	RMSEA[a] (90% CI)	SRMR[b]	CFI[c]	TLI[d]	AIC[e]
1-Factor Model	577.67 (90)	<.001	.14 (0.13-0.15)	.13	.88	.89	13135.05
2-Factor Model	263.16 (79)	<.001	.09 (0.08-0.10)	.09	.96	.95	12842.55
3-Factor Model	138.95 (67)	<.001	.06 (0.05-0.08)	.08	.98	.98	12742.33
4-Factor Model	108.12 (54)	<.001	.06 (0.04-0.08)	.08	.99	.98	12737.50

[a]RMSEA: root mean square error of approximation.

[b]SRMR: standardized root mean square residual.

[c]CFI: comparative fit index.

[d]TLI: Tucker-Lewis index.

[e]AIC: Akaike information criterion.

Although the 4-factor model yielded the best fitting model, items from the scale did not statistically significantly load onto the fourth factor. Therefore, the 3-factor model was used as the basis for assessing measurement invariance among young and old respondents.

Table 3 shows the statistically significant unstandardized factor loadings for the 3-factor model among Millennials, Generation X, and Baby Boomers/Silent Generation groups. Factor 1, which includes items that assess awareness about what health information is available on the Internet and where it can be located, contained significant factor loadings for Items 1-2 across all groups. Similarly, items 5-8 yielded high (greater than .40) and significant loadings on Factor 3, which included items that assess confidence in evaluating and using health information to answer health-related questions. Items 3 and 4, which assessed knowledge about how to use and find helpful health resource on the Internet, had a moderate to strong relationship with Factor 2 across all generation groups. All 3 factors were statistically significantly correlated with one another across all 3 groups. Interestingly, the correlation of Factor 1 with Factors 2 (r=.98) and 3 (r=.80) were much stronger than for the other generation groups. The final 3-factor model used to guide measurement invariance testing is shown in Figure 1.

Table 3. The 3-factor loadings for each generation.

Electronic Health Literacy Scale Item	Millennials[a]			Generation X[b]			Baby Boomer/Silent Generation[c]		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
(E1) I know what health resources are available on the Internet	.69[d]	.01	.06	.69[d]	.01	.06	.69[d]	.01	.06
(E2) I know where to find helpful health resources on the Internet	.64[d]	.18[e]	–.01	.64[d]	.18[e]	–.01	.64[d]	.18[e]	–.01
(E3) I know how to use the health information I find on the Internet to help me	.07	.39[f]	.22[f]	.07	.40[f]	.22[f]	.07	.40[f]	.22[f]
(E4) I know how to find helpful health resources on the Internet	–.02	.78[d]	.01	–.02	.78[d]	.01	–.02	.78[d]	.01
(E5) I have the skills I need to evaluate the health resources I find on the Internet.	–.09	.02	.72[d]	–.09	.02	.72[d]	–.09	.02	.72[d]
(E6) I know how to use the Internet to answer my questions about health.	–.01	.18[d]	.45[d]	–.01	.18[d]	.45[d]	–.01	.18[e]	.45[d]
(E7) I can tell high quality health resources from low quality health resources on the Internet	.17	–.03	.49[f]	.17	–.03	.49[d]	.17	–.03	.50[d]

[a]Factor 1 with Factor 2 (r=.70, P<.001), Factor 1 with Factor 3 (r=.63, P<.001), Factor 2 with Factor 3 (r=.76, P<.001).

[b]Factor 1 with Factor 2 (r=.98, P<.001), Factor 1 with Factor 3 (r=.80, P<.001), Factor 2 with Factor 3 (r=.77, P<.001).

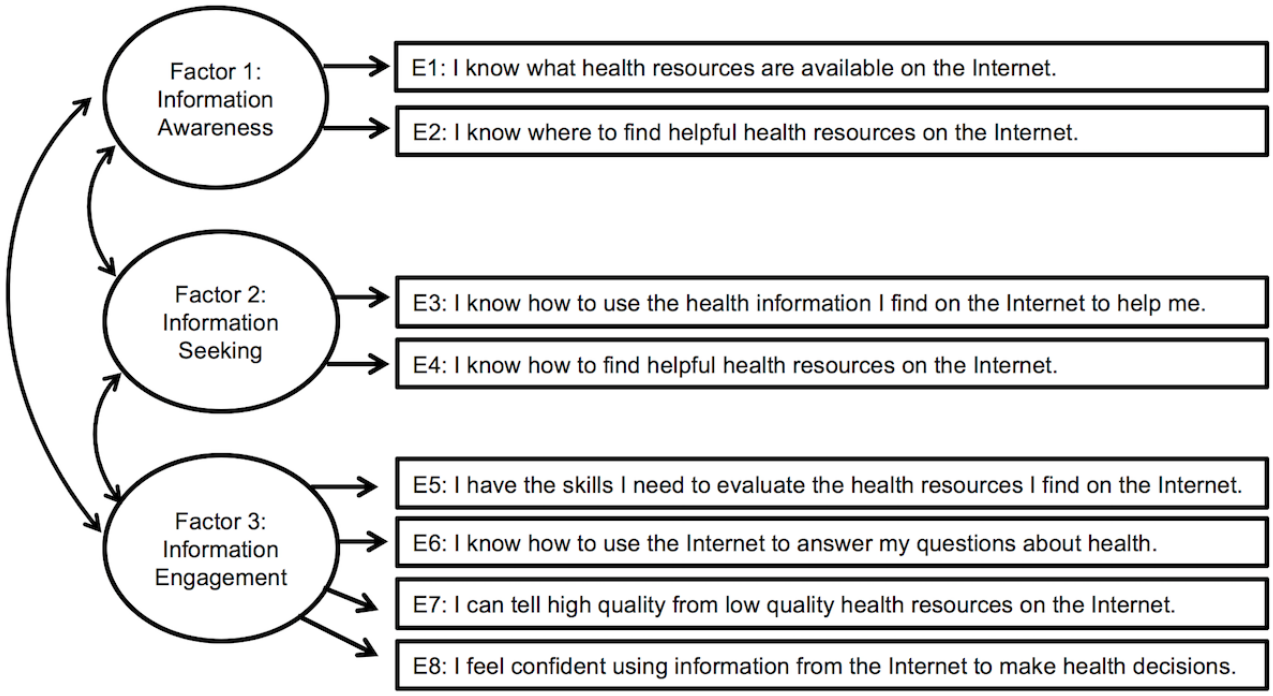
[c]Factor 1 with Factor 2 (r=.79, P<.001), Factor 1 with Factor 3 (r=.79, P<.001), Factor 2 with Factor 3 (r=.89, P<.001).

[d]P<.001

[e]P<.05

[f]P<.01

Figure 1. Proposed 3-factor electronic health literacy scale (eHEALS) measurement model.



Measurement Invariance

Table 4 shows the results for configural, pattern, and unique factor invariance tests through the use of a CFA. The 3-factor model has slightly poorer, but acceptable, model fit in regards to configural invariance. This is determined from the RMSEA (value=.08, 90% CI 0.06-0.09) and CFI/TLI (.98 and .97, respectively), confirming that the 3-factor model represents the factor structure of eHEALS across all generations. Adding constraints on the factor loadings across generation groups (pattern invariance testing) resulted in slight improvement of RMSEA and relatively steady SRMR, CFI, and TLI values. The change in chi-square was not statistically significant and the CFI did not deviate by .01. In regard to unique factor invariance, the change in chi-square ($\Delta\chi^2$=69.51, Δdf=16) was statistically significant at. P<.05. As such, unique factor invariance was rejected as equating the error variances of each item across groups significantly diminished the model fit. Moreover, the AIC value for the pattern invariance model (value=12,770.60) was lower than the models testing for configural (value=12,775.72) and unique factor (value=12,808.11) invariance. Therefore, measurement invariance of for the proposed 3-factor structure exists among Millennials, the Generation X, and Baby Boomers/Silent Generation.

Electronic Health Literacy Scale Scores by Age Group

Table 5 shows the average scale scores for the 3-factor eHEALS model across each generation. Internal consistency alpha estimates were within appropriate range for each factor, and omega coefficients demonstrated equivalent values to support reliability of the data. A one-way ANOVA showed that eHEALS scores varied across generations for Factor 1 (F [2, 827]=8.17, P<.001), Factor 2 (F [2, 826]=6.00, P=.003), and Factor 2 (F [2, 827]=18.51, P<.001). Tukey honest significant difference (HSD) post hoc analyses showed that, on average, members of the Baby Boomer and Silent Generation groups reported less knowledge and confidence in their eHealth literacy across all factors (P<.05), as compared to members of the Millennial and Generation X groups.

Table 4. Fit statistic summary for testing measurement invariance in the 3-factor model of electronic health literacy scale.

Model	χ^2 (df)	RMSEA[a]	SRMR[b]	CFI[c]	TLI[d]	AIC[e]	Model comparison, $\Delta\chi^2$ (Δdf)
Model 1: Configural Invariance	160.33 (61)	.08	.03	.98	.97	12775.72	0.0 (0)
Model 2: Pattern Invariance	175.21 (71)	.07	.05	.98	.97	12770.60	14.88 (10)
Model 3: Unique Factor Invariance	244.72 (87)	.08	.08	.96	.96	12808.11	69.51[f] (16)

[a]RMSEA: root mean square error of approximation.

[b]SRMR: standardized root mean square residual.

[c]CFI: comparative fit index.

[d]TLI: Tucker-Lewis index.

[e]AIC: Akaike Information Criterion.

[f]P<.05.

Table 5. Average eHealth literacy scores by age group.

Electronic Health Literacy Scale Factor	Millennials			Generation X			Baby Boomer/Silent Generation			Total		
	α	ω	Mean (SD)	α	ω	Mean (SD)	α	ω	Mean (SD)[a]	α	ω	Mean (SD)
Factor 1: Information Awareness[b]	.80	.80	7.69 (1.68)	.91	.91	7.72 (1.80)	.83	.83	7.22 (1.67)	.84	.84	7.48 (1.71)
Factor 2: Information Seeking[c]	.86	.86	8.02 (1.50)	.90	.91	8.02 (1.57)	.89	.89	7.66 (1.46)	.88	.88	7.85 (1.51)
Factor 3: Information Engagement[d]	.79	.79	15.37 (2.66)	.85	.84	15.55 (2.77)	.86	.86	14.25 (2.97)	.84	.84	14.89 (2.88)

[a]$P<.05$.

[b]Factor 1 (min score=2; max score=10).

[c]Factor 2 (min score=2; max score=10).

[d]Factor 3 (min score=4; max score=20).

Discussion

Principal Findings

This study examined the degree of measurement invariance in eHEALS scores in the United States belonging to the Millennial, Generation X, and Baby Boomers/Silent Generations. The eHEALS is a multidimensional measure that can be used to assess eHealth literacy across the lifespan consistently. Millennials are more knowledgeable and confident in their online health information awareness, information seeking skills, and information engagement abilities, as compared to members of Generation X and the Baby Boomers/Silent Generation. Further, this study offers significant implications for the continued use and potential refinement of eHEALS in future research and practice-based settings.

The eHEALS scores best fit a positively correlated 3-factor model that captures the following underlying factors: *information awareness*, *information seeking*, and *information engagement*. This finding comes at a time when there is inconsistent evidence for the factor structure of eHEALS. Results of our study contrast with those described by Nguyen and colleagues [52], who explored the dimensionality of eHEALS when it was administered online to a significant proportion (60%) Millennials. Data from Nguyen and colleagues [52] showed eHEALS to have a unidimensional structure with a principal components analysis, which traditionally identifies the fewest number of factors that explain the substantial amount of variance in observed variables [78]. Considering the conflicting evidence describing the dimensionality of eHealth literacy, our alternative multi-group exploratory structural equation modeling approach sought to validate constructs implicit within eHEALS items across three different age groups. Moreover, the current study strived to cast a broader net to explore not only which eHEALS items best explained retained factors, but also how these factors might function in a theoretically driven manner consistent with eHealth literacy literature. Contrary to findings reported by Nguyen and colleagues [52], evidence generated in this study supported a 3-factor model of the English-version of eHEALS. Of note, our sample of Web-based panelists included proportionately more adults representing older Baby Boomer and Silent Generations. These 2 generations were underrepresented in their analyses conducted using an Internet-based sample obtained through machine learning software.

The 3-factor eHEALS model supported in this study captures a more precise assessment of eHealth literacy that goes beyond individual knowledge and perceptions of behavioral capability. The 3-factor eHEALS model comprises items that measure self-efficacy towards central operational skills related to eHealth literacy (ie, locate, evaluate, apply). These operational skills are associated with unique, albeit related, dimensions of self-efficacy in the context of eHealth literacy [16], which explains the highly correlated 3-factor model containing unique factors.

Configural and pattern invariance was upheld across all generation groups in the 3-factor eHEALS model, suggesting that eHEALS scores from the 3-factor model can be interpreted equivalently, regardless of respondents' age group membership. Despite invariance between groups in the current study, the items that comprise each of these factors are inconsistent with the results of previous literature. Stellefson and colleagues [55] examined the factor structure of eHEALS scores among Baby Boomers and Silent Generation members during a telephone interview and found that Item 3 (ie, "I have the skills I need to evaluate the health resources I find on the Internet") significantly loaded on both Factor 2 and Factor 3. In that study, Factor 2 included items related to knowing how to use and find helpful health information on the Internet to make informed health decisions. Factor 3 included only 1 other item with a significant factor loading, which addressed the ability to evaluate the quality of online health information. Findings from Stellefson and colleagues [55] run contrary to the current study and also findings reported by Sudbury-Riley and colleagues' [49], who speculated that the content and theoretical underpinnings of this particular eHEALS item (Item 3) denote skills related to confidence in the ability to evaluate and act upon health information from the Internet. After a closer inspection of Item 3 content, it appears that this question may assess two distinct skills: (1) can one evaluate health information from the Internet? and (2) can one find health information on the Internet? The mode of data collection in Stellefson and colleagues' [55] study was over the telephone, whereas the data collected in the current study and Sudbury-Riley and colleagues' [49] was through a Web-based survey. It is possible that respondents only cognitively processed a single operational behavior outlined in

this item (ie, find, evaluate), or perhaps the telephone interviewer placed emphasis on one skill over the other. Future research is needed to understand how data collection modality (eg, telephone, online) might directly affect the interpretation of the eHEALS items and ultimately the construct validity of the data produced.

Lastly, the final test of measurement invariance proved to be insufficient. The residual error variances of items were significantly different across age groups when tested within the 3-factor eHEALS model. Unique factor invariance is the strictest form of measurement equivalence. It is rarely achieved in practice, and experts have recently acknowledged that establishing unique factor invariance can be somewhat unreasonable for subjective measurement [79]. Therefore, we suggest that scores produced by the eHEALS may still be used as a comparative index to examine eHealth literacy across age groups [72].

Limitations

This study sampled opt-in respondents from a Qualtrics Survey Panel taken from the general US population. Despite the population from which the sample was derived, the respondents were predominantly female with a normally distributed income and educational level. Moreover, half of the sample identified as Caucasian and the other half as Black/African American. In other words, this study enrolled over 400 respondents from population subgroups (ie, middle-older age adults, Black/African Americans) that are traditionally underserved in health promotion research. Although this represents a limitation affecting the generalizability of data to the entire US population, the diversity of sample characteristics remains a significant strength of this measurement study.

This was a self-reported Web-based survey, and, therefore, the results of this study can only speak to the interpretation and measurement invariance of scores from eHEALS administered on the Web. There is sufficient reliability and validity evidence of eHEALS delivered via telephone among middle-to-older age adults [55], a population most likely to respond differently to Web-based versus telephone-administered surveys [80]. Future research could explore the degree of measurement invariance of the 3-factor eHEALS model across generations according to the mode of survey administration. Moreover, respondents of this Web-based survey were members of Qualtrics Panels who opted-in to participate. The purposefully racially stratified sample and normally distributed income levels compromises the generalizability of the findings. However, the oversampling of minorities and low-income adults engaged these particularly vulnerable and hard-to-reach populations in survey research.

Although this study did not consider the geographic region (ie, rural versus urban) of the sample, nearly 70% of the sample reported using social media for health-related purposes, which requires a sufficient level of broadband. Rural adults are generally older [81] and have limited broadband connections [82] that enable sustained access to eHealth services. Moreover, rural residents are nearly twice as likely to not use the Internet as compared to their urban counterparts [82,83]. Therefore, factors beyond geographic location may limit rural adults' eHealth use. Based on the limited empirical evidence related to the eHealth literacy of rural populations [84], future research is needed to explore eHealth literacy and its measurement among populations according to rurality regarding physical space (ie, Rural-Urban Commuting Area or Metropolitan Statistical Area data) and sociocultural rural identity.

Practical Implications

Acknowledging the multidimensionality of scores obtained from eHEALS will allow practitioners to obtain a more precise understanding of consumers strengths and weaknesses using the Internet for health-related purposes. Rather than ambiguously interpreting "low eHealth literacy" based on prior unidimensional assumptions underlying eHEALS, practitioners considering the 3-factor model of eHEALS can identify the degree to which their patients have confidence in online health information awareness, search, and engagement. Interpreting scores based on 3 underlying eHEALS dimensions can assist practitioners and researchers to more efficiently direct patients to eHealth resources that are appropriate to their relative skill set, whether it is simply increasing awareness of existing online health information resources or providing a direct link to a particular website with credible health information. Precisely identifying limitations in core operational behaviors central to eHealth literacy will help to inform more tailored and efficient eHealth literacy interventions that consider an individual's perceptions of technology adoption and acceptability.

Compared with Millennials, older generations reported lower knowledge and self-efficacy in each of the factors captured by eHEALS. Specifically, adults belonging to Generation X and the Baby Boomers/Silent Generation had less confidence in their (1) awareness of online health information, (2) skills to locate online health information, and (3) ability to evaluate and act on health information once it is located online. This finding is consistent with previous literature stating that older adults have lower proficiency in eHealth literacy than their younger counterparts [18,22,25,26]. However, it is currently difficult to measure the degree to which specific eHealth literacy skills are deficient across different age segments. Our study helps to shed light on how to interpret eHEALS scores, such that information is gathered regarding which particular eHealth literacy skills are limited and the degree to which they are limited across age groups. eHEALS has strong potential to be used as the standard assessment tool for coordinating eHealth literacy training interventions based on these three discrete factors. For example, structured interventions could be delivered in three modules where skill-building activities aim to improve eHealth awareness, as well as information seeking and evaluation. Although older and younger adults respond differently to eHealth literacy interventions [85], these 3 factors (ie, skill sets) are central components of eHealth literacy, and thus should be considered in the planning, implementation, and evaluation of training interventions designed to improve the eHealth literacy of older adults through narrowing the chasm that currently exists between eHealth adoption and sustained use.

Finally, the results of this study provide implications for refining and updating the eHEALS. The brevity of eHEALS makes it an ideal scale for use in research and clinical care. However, it is necessary to ensure that there is an adequate number of items

that correspond to each factor. Some measurement guidelines support the reliability of highly correlated factors that only comprise 2 items each [86,87]. However, other measurement standards recommend including at least 3 items per factor [88]. In the current study, the 3 eHEALS factors were correlated to a statistically significant degree. The strong factorial relationship allowed the model to function adequately with fewer items on Factors 1 and 2. This finding is contrary to the findings reported by Sudbury-Riley and colleagues [49], who found that only 1 latent factor (ie, online health information awareness) was best reflected by 2 eHEALS items, whereas factors related to information seeking and application (eg, knowing how to find and use online health information, self-efficacy to evaluate, and use online health information) were comprised of 3 items each. Further research is needed to develop unbiased items that sufficiently capture the theoretical underpinnings of eHealth literacy and its multidimensional constructs. Moreover, to account for the dynamic and interactive nature of eHealth [2], future research can build upon our findings to create and test new items that account for a fourth latent factor that captures "social" skills related to eHealth literacy.

Conclusion

Valid age group comparisons can be made with the 3-factor structure of eHEALS among Millennials, Generation Xers, and Baby Boomer/Silent Generation members. Results of this study add to the library of literature showing that older adults have significantly lower eHealth literacy scores as compared to younger adults. Specifically, this study supports that members of younger generations have a greater awareness of eHealth resources and more confidence in their information seeking and engagement skills on the Internet, as compared to older generations. The brevity of eHEALS coupled with its multi-dimensional structure can assist health care practitioners and researchers in tailoring eHealth literacy interventions designed to augment user performance on these relevant constructs. Furthermore, findings of this study have significant implications for more precisely measuring and improving eHealth literacy skills across the lifespan.

Acknowledgments

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH) under Award Number F31HL132463. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Mars M, Scott RE. Global e-health policy: a work in progress. Health Aff (Millwood) 2010 Feb;29(2):237-243. [doi: 10.1377/hlthaff.2009.0945] [Medline: 20348067]
2. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a Conceptual Model for eHealth: Qualitative Study with Key Informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
3. Pew Research Center. 2018. Internet/Broadband Fact Sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [WebCite Cache ID 6xwoPoPtt]
4. Anderson M, Perrin A. Pew Research Center. 2017. Technology Use Among Seniors URL: http://www.pewinternet.org/2017/05/17/technology-use-among-seniors/ [WebCite Cache ID 6zDu4YMFt]
5. Smith A. Pew Research Center. 2014. Older Adults and Technology Use URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [WebCite Cache ID 6P34DTel4]
6. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of Technology: Frequency of Use for Younger and Older Adults. Ageing Int 2011 Mar;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
7. Chung JE, Park N, Wang H, Fulk J, McLaughlin M. Age differences in perceptions of online community participation among non-users: An extension of the Technology Acceptance Model. Computers in Human Behavior 2010;26(6):1674-1684.
8. Vaportzis E, Clausen MG, Gow AJ. Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
9. Chiu C, Liu C. Understanding Older Adult's Technology Adoption and Withdrawal for Elderly Care and Education: Mixed Method Analysis from National Survey. J Med Internet Res 2017 Nov 03;19(11):e374 [FREE Full text] [doi: 10.2196/jmir.7401] [Medline: 29101093]
10. Anderson M, Perrin A. Pew Research Center. 2017. Barriers to adoption and attitudes towards technology URL: http://www.pewinternet.org/2017/05/17/barriers-to-adoption-and-attitudes-towards-technology/ [WebCite Cache ID 6zDveEQ11]
11. Young R, Willis E, Cameron G, Geana M. "Willing but unwilling": attitudinal barriers to adoption of home-based health information technology among older adults. Health Informatics J 2014 Jun;20(2):127-135. [doi: 10.1177/1460458213486906] [Medline: 24056750]

12. Choi N. Relationship between health service use and health information technology use among older adults: analysis of the US National Health Interview Survey. J Med Internet Res 2011 Apr 20;13(2):e33 [FREE Full text] [doi: 10.2196/jmir.1753] [Medline: 21752784]
13. Decker SL, Jamoom EW, Sisk JE. Physicians in nonprimary care and small practices and those age 55 and older lag in adopting electronic health record systems. Health Aff (Millwood) 2012 May;31(5):1108-1114. [doi: 10.1377/hlthaff.2011.1121] [Medline: 22535502]
14. MORRIS MG, VENKATESH V. AGE DIFFERENCES IN TECHNOLOGY ADOPTION DECISIONS: IMPLICATIONS FOR A CHANGING WORK FORCE. Personnel Psychology 2000 Jun;53(2):375-403. [doi: 10.1111/j.1744-6570.2000.tb00206.x]
15. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008 Jan;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
16. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
17. Paige SR, Krieger JL, Stellefson ML. The Influence of eHealth Literacy on Perceived Trust in Online Health Communication Channels and Sources. J Health Commun 2017 Jan;22(1):53-65. [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
18. 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]
19. Werts N, Hutton-Rogers L. American Journal of Health Sciences. 2013. Barriers to achieving e-health literacy URL: https://clutejournals.com/index.php/AJHS/article/view/8007 [accessed 2018-06-26] [WebCite Cache ID 70SsBFU6i]
20. Milne RA, Puts MTE, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of High eHealth Literacy in Primary Lung Cancer Survivors. J Cancer Educ 2015 Dec;30(4):685-692. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]
21. Rios GR. eHealth Literacy and Older Adults. Topics in Geriatric Rehabilitation 2013;29(2):116-125. [doi: 10.1097/TGR.0b013e31827ec0bd]
22. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. J Med Internet Res 2016 Jul;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
23. Gilstad H. Toward a comprehensive model of eHealth literacy. URL: http://ceur-ws.org/Vol-1251/paper7.pdf [WebCite Cache ID 6xKLH9H1V]
24. Bautista JR. From solving a health problem to achieving quality of life: redefining eHealth literacy. Journal of Literacy and Technology 2015;16(2):33-54.
25. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
26. 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;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
27. Chesser AK, Keene Woods N, Smothers K, Rogers N. Health Literacy and Older Adults. Gerontology and Geriatric Medicine 2016 Mar 15;2:233372141663049. [doi: 10.1177/2333721416630492] [Medline: 28138488]
28. Speros CI. More than words: Promoting health literacy in older adults. Online J Issues Nurs 2009;14(3):5. [doi: 10.3912/OJIN.Vol14No03Man05]
29. Bennett JA, Flaherty-Robb M. Issues affecting the health of older citizens: Meeting the challenge. Online J Issues Nurs 2003;8(2):2. [Medline: 12795628]
30. US Department of Health and Human Services. 2018. Older adults: Healthy People 2020 URL: https://www.healthypeople.gov/2020/topics-objectives/topic/older-adults [WebCite Cache ID 6xwoZm85C]
31. Michaels Fisher H, McCabe S. Managing chronic conditions for elderly adults: The VNS CHOICE Model. Health Care Financ Rev 2005;27(1):33-45. [Medline: 17288076]
32. 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/ [WebCite Cache ID 6xwofPsqO]
33. McCray AT. Promoting health literacy. J Am Med Inform Assoc 2005 Nov;12(2):152-163 [FREE Full text] [doi: 10.1197/jamia.M1687] [Medline: 15561782]
34. Kaye LW. Older adults, rural living, and the escalating risk of social isolation. Public Policy & Aging Report 2018;27(4):139-144. [doi: 10.1093/ppar/prx029]
35. Valtorta N, Hanratty B. Loneliness, isolation and the health of older adults: do we need a new research agenda? J R Soc Med 2012 Dec;105(12):518-522 [FREE Full text] [doi: 10.1258/jrsm.2012.120128] [Medline: 23288086]
36. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
37. Connaway LS, Radford M, Dickey T, De A, Confer P. Sense-making and synchronicity: Information-seeking behaviors of Millennials and Baby Boomers. International Journal of Libraries and Information Studies 2008;58(2):123-135.

38. Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. Exp Aging Res 1992;18(3-4):117-144. [doi: 10.1080/03610739208253916] [Medline: 1459160]
39. Vandenberg RJ, Lance CE. A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. Organizational Research Methods 2016 Jun 29;3(1):4-70. [doi: 10.1177/109442810031002]
40. Tran TV. Testing Cross-Cultural Measurement Invariance. Developing Cross Cultural Measurement 2009:131. [doi: 10.1093/acprof:oso/9780195325089.003.0007]
41. Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. Psychol Bull 1993 Nov;114(3):552-566. [Medline: 8272470]
42. Harada CN, Natelson LMC, Triebel KL. Normal cognitive aging. Clin Geriatr Med 2013 Nov;29(4):737-752 [FREE Full text] [doi: 10.1016/j.cger.2013.07.002] [Medline: 24094294]
43. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
44. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an Informed Consumer of Health Information and Assessment of Electronic Health Literacy in a National Sample of Internet Users: Validity and Reliability of the e-HLS Instrument. J Med Internet Res 2016 Jul 11;18(7):e161 [FREE Full text] [doi: 10.2196/jmir.5496] [Medline: 27400726]
45. van DVR, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
46. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014;15:18 [FREE Full text] [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]
47. Griebel L, Enwald H, Gilstad H, Pohl A, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2017 Oct 18:1-16. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
48. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
49. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
50. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A Multidimensional Tool Based on the eHealth Literacy Framework: Development and Initial Validity Testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
51. 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]
52. Nguyen J, Moorhouse M, Curbow B, Christie J, Walsh-Childers K, Islam S. Construct Validity of the eHealth Literacy Scale (eHEALS) Among Two Adult Populations: A Rasch Analysis. JMIR Public Health Surveill 2016 May;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]
53. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: An item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb;100(2):320-326. [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
54. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156. [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
55. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and Validity of the Telephone-Based eHealth Literacy Scale Among Older Adults: Cross-Sectional Survey. J Med Internet Res 2017 Oct 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
56. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
57. Soellner R, Huber S, Reder M. The Concept of eHealth Literacy and Its Measurement. Journal of Media Psychology 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
58. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. The European Health Psychologist 2015;17(6):275-280.
59. Hyde LL, Boyes AW, Evans T, Mackenzie LJ, Sanson-Fisher R. Three-Factor Structure of the eHealth Literacy Scale Among Magnetic Resonance Imaging and Computed Tomography Outpatients: A Confirmatory Factor Analysis. JMIR Hum Factors 2018 Feb 19;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
60. Pew Research Center. 2015. The Whys and Hows of Generations Research URL: http://www.people-press.org/2015/09/03/the-whys-and-hows-of-generations-research/ [WebCite Cache ID 6zDwTEwWe]
61. No authors listed. Clarification to Brook and Willoughby (2016). Dev Psychol 2016 Dec;52(7):1164-1167. [doi: 10.1037/dev0000177] [Medline: 27337517]

62. Pilkington R, Taylor AW, Hugo G, Wittert G. Are Baby Boomers healthier than Generation X? A profile of Australia's working generations using National Health Survey data. PLoS One 2014 Mar;9(3):e93087 [FREE Full text] [doi: 10.1371/journal.pone.0093087] [Medline: 24671114]
63. Fronstin P, Elmlinger A. Consumer engagement in health care and choice of health plan: Differences among Millennials, Baby Boomers, and Generation X have implications for plan sponsors. Rochester, NY: Social Science Research Network; 2017.
64. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
65. van DVR, van DAJ, Drossaert CH, Taal E, van DJA, van DLMA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
66. Health Information National Trends Survey. URL: http://hints.cancer.gov/Default.aspx [accessed 2013-11-08] [WebCite Cache ID 6Kz4grSh7]
67. Fry R. Pew Research Center. 2018. Millennials projected to overtake Baby Boomers as America's largest generation URL: http://www.pewresearch.org/fact-tank/2018/03/01/millennials-overtake-baby-boomers/ [WebCite Cache ID 6ycrfV9xj]
68. Muthen LK, Muthen BO. Mplus: Statistical analysis with latent variables. 2015. URL: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf [WebCite Cache ID 6xwojzA9U]
69. Asparouhov T, Muthén B. Exploratory Structural Equation Modeling. Structural Equation Modeling: A Multidisciplinary Journal 2009 Jul 14;16(3):397-438. [doi: 10.1080/10705510903008204]
70. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;5(1):1-55.
71. Costello AB, Osborne J. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation 2005;10(7):1-9.
72. van de Schoot R, Lugtig P, Hox J. A checklist for testing measurement invariance. European Journal of Developmental Psychology 2012 Jul;9(4):486-492. [doi: 10.1080/17405629.2012.686740]
73. Schumacker RE, Lomax RG. A Beginner's Guide to Structural Equation Modeling: Third Edition. New York: Routledge; 2010.
74. Cheung GW, Rensvold RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal 2002 Apr;9(2):233-255. [doi: 10.1207/S15328007SEM0902_5]
75. IBM Corp. SPSS Statistics 24 - United States URL: http://www-01.ibm.com/support/docview.wss?uid=swg24041224 [WebCite Cache ID 6xwom6VtR]
76. Trizano-Hermosilla I, Alvarado JM. Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. Front Psychol 2016;7:769 [FREE Full text] [doi: 10.3389/fpsyg.2016.00769] [Medline: 27303333]
77. Deng L, Chan W. Testing the Difference Between Reliability Coefficients Alpha and Omega. Educational and Psychological Measurement 2016 Jul 20;77(2):185-203. [doi: 10.1177/0013164416658325]
78. Stellefson ML, Hanik BW, Chaney BH, Chaney JD. Factor retention in EFA: strategies for health behavior researchers. Am J Health Behav 2009;33(5):587-599. [Medline: 19296749]
79. Bialosiewicz S, Murphy K, Berry T. Do our measures measure up? The critical role of measurement invariance. Claremont, CA: Claremont Evaluation Center; 2013.
80. Dillman DA, Smyth JD, Christian LM. Internet, mail,mixed-mode surveys: The tailored design method. 3rd ed. Hoboken: Wiley; 2008.
81. Day JC, Hays D, Smith A. United States Census Bureau. 2016. A glance at the age structure and labor force participation of rural America URL: https://www.census.gov/newsroom/blogs/random-samplings/2016/12/a_glance_at_the_age.html [WebCite Cache ID 6xwoscBQv]
82. Perrin A. Pew Research Center. 2017. Digital gap between rural and nonrural America persists URL: http://www.pewresearch.org/fact-tank/2017/05/19/digital-gap-between-rural-and-nonrural-america-persists/ [WebCite Cache ID 6xwow0Qay]
83. Anderson M, Perrin A, Jiang J. Pew Research Center. 2018. 11% of Americans don't use the internet URL: http://www.pewresearch.org/fact-tank/2018/03/05/some-americans-dont-use-the-internet-who-are-they/ [WebCite Cache ID 6xwp1Iwa6]
84. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: A systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016 Feb;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
85. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
86. Worthington RL, Whittaker TA. Scale Development Research. The Counseling Psychologist 2016 Jun 30;34(6):806-838. [doi: 10.1177/0011000006288127]
87. Yong AG, Pearce S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in Quantitative Methods for Psychology 2013;9(2):79-94.

88. Raubenheimer J. An item selection procedure to maximise scale reliability and validity. SA J Ind Psychol 2004 Oct 26;30(4):59-64. [doi: 10.4102/sajip.v30i4.168]

Abbreviations

AIC: Akaike information criterion
CFA: confirmatory factor analysis
CFI: comparative fix index
eHEALS: electronic health literacy scale
IRB: Institutional Review Board
MG-ESEM: multi-group exploratory structural equation models
NIH: National Institutes of Health
NHLBI: National Heart, Lung, and Blood Institute
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 21.03.18; peer-reviewed by L Sudbury-Riley, E Chavarria, B Chaney, B Curbow; comments to author 09.04.18; revised version received 17.05.18; accepted 16.06.18; published 09.07.18.

Please cite as:
Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J
Electronic Health Literacy Across the Lifespan: Measurement Invariance Study
J Med Internet Res 2018;20(7):e10434
URL: http://www.jmir.org/2018/7/e10434/
doi:10.2196/10434
PMID:29986848

Original Paper

Evaluating Doctor Performance: Ordinal Regression-Based Approach

Yong Shi[1,2,3,4], PhD; Peijia Li[5,6], PhD; Xiaodan Yu[7], PhD; Huadong Wang[2,3], PhD; Lingfeng Niu[1,2,3], PhD

[1]School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China

[2]Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing, China

[3]Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, China

[4]College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE, United States

[5]School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing, China

[6]Development and Planning Research Institute, Ningbo, China

[7]School of Information Technology and Management, University of International Business and Economics, Beijing, China

Corresponding Author:
Lingfeng Niu, PhD
Research Center on Fictitious Economy and Data Science
Chinese Academy of Sciences
80 Zhongguancun East Road
Haidian District
Beijing, 100190
China
Phone: 86 15600616246
Email: niulf@ucas.ac.cn

Abstract

Background: Doctor's performance evaluation is an important task in mobile health (mHealth), which aims to evaluate the overall quality of online diagnosis and patient outcomes so that customer satisfaction and loyalty can be attained. However, most patients tend not to rate doctors' performance, therefore, it is imperative to develop a model to make doctor's performance evaluation automatic. When evaluating doctors' performance, we rate it into a score label that is as close as possible to the true one.

Objective: This study aims to perform automatic doctor's performance evaluation from online textual consultations between doctors and patients by way of a novel machine learning method.

Methods: We propose a solution that models doctor's performance evaluation as an ordinal regression problem. In doing so, a support vector machine combined with an ordinal partitioning model (SVMOP), along with an innovative predictive function will be developed to capture the hidden preferences of the ordering labels over doctor's performance evaluation. When engineering the basic text features, eight customized features (extracted from over 70,000 medical entries) were added and further boosted by the Gradient Boosting Decision Tree algorithm.

Results: Real data sets from one of the largest mobile doctor/patient communication platforms in China are used in our study. Statistically, 64% of data on mHealth platforms lack the evaluation labels from patients. Experimental results reveal that our approach can support an automatic doctor performance evaluation. Compared with other auto-evaluation models, SVMOP improves mean absolute error (MAE) by 0.1, mean square error (MSE) by 0.5, pairwise accuracy (PAcc) by 5%; the suggested customized features improve MAE by 0.1, MSE by 0.2, PAcc by 3%. After boosting, performance is further improved. Based on SVMOP, predictive features like politeness and sentiment words can be mined, which can be further applied to guide the development of mHealth platforms.

Conclusions: The initial modelling of doctor performance evaluation is an ordinal regression problem. Experiments show that the performance of our proposed model with revised prediction function is better than many other machine learning methods on MAE, MSE, as well as PAcc. With this model, the mHealth platform could not only make an online auto-evaluation of physician performance, but also obtain the most effective features, thereby guiding physician performance and the development of mHealth platforms.

(J Med Internet Res 2018;20(7):e240) doi:10.2196/jmir.9300

KEYWORDS

performance evaluation; ordinal regression; mHealth; support vector machines; ordinal partitioning

Introduction

With the advancement of the internet and electronic devices, mobile heath (mHealth), is defined by the World Health Organization as "medical and public health practice supported by mobile devices," is becoming increasingly popular. mHealth has strong links with electronic health [1] with some differences [2]. According to an mHealth survey [3], 80% of physicians use smartphones and medical apps and 61% of people have downloaded a medical app. Meanwhile, 93% of physicians believe that mHealth apps can help to improve patients' health. The doctor/patient communication platform is one of the most common areas in mHealth, for example, "Dermatologist-on-Call" in America and, in China, "Chunyu-Doctor-online" and "Good-Doctor-online." These platforms digitally connect doctors and patients and offer a convenient channel for doctor/patient communication and help doctors use time more efficiently. Additionally, the mHealth platforms are more beneficial to under-developed countries, especially when medical resources are scarce, and quality medical care is difficult to access.

Many doctor/patient communication platforms face the challenge of how to evaluate the performance of doctors online. Doctor performance evaluation serves to increase the probability for patients to have a positive experience and improve patient satisfaction [4-6]. Meanwhile, doctor performance evaluation also helps doctors to improve medical practice [7]. In this paper, we address the issue of doctor performance evaluation (DPE).

Various methods have been attempted that address the issue of DPEs. Ratings by patients is the most common method, which averages patient ratings when evaluating physicians. Physician ratings are usually based on the following labels: (1) very unsatisfied, (2) unsatisfied, (3) neutral, (4) satisfied, and (5) very satisfied. Statistics show that only a small proportion of patients rate their doctors on mHealth platforms, and in China, only 36% of patients rated their doctor at the end of the consultation.

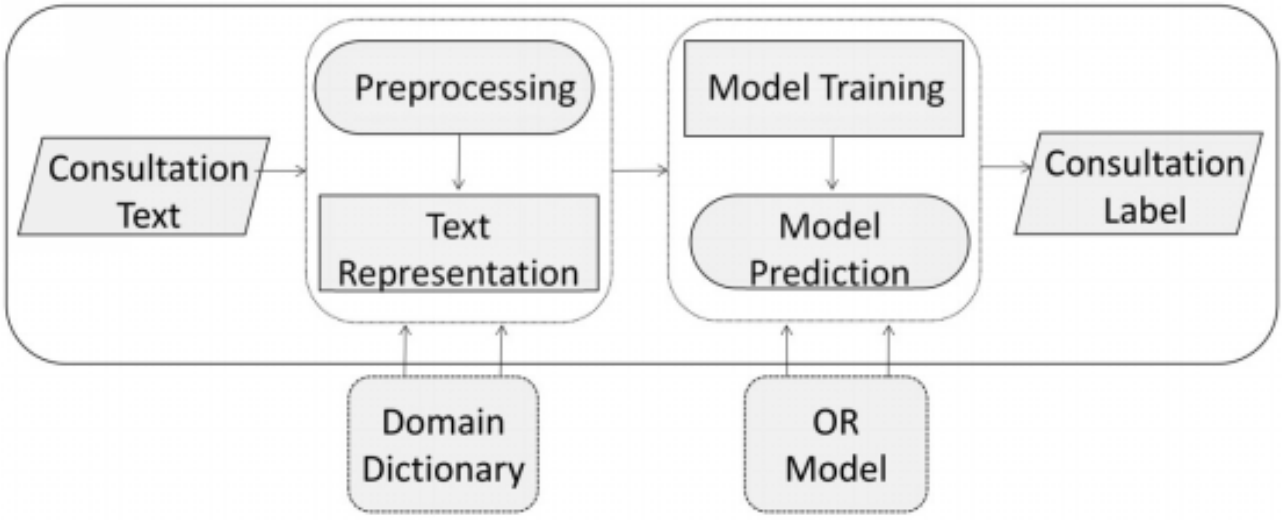
A physician expert assesses the doctors' professional skills and services. In this combined method, experts re-rate the patient's unsatisfied consultations and judge whether the doctors are qualified. It is an advanced evaluation method, which not only considers patient ratings but also imbues prior professional knowledge. Therefore, this method is recommended but heavily depends on the patient's ratings.

Considering the amount of data generated from doctor/patient communication platforms every day, machine learning techniques are recommended. In the machine learning area, some scholars have rated patient satisfaction into standard classification algorithms [8,9] but ignore the ordering information between labels. The ordering information between labels as mentioned above are, (1) very unsatisfied, (2) unsatisfied, (3) neutral, (4) satisfied, and (5) very satisfied. The label "unsatisfied" is adjacent to the label "very unsatisfied" and the label "neutral," while the label "very unsatisfied" is not next to the label "neutral," therefore the ordering of information is extremely important. Rating a "very unsatisfied" doctor consultation as an "unsatisfied" one is less of an error than rating it as "very satisfied." Therefore, it is important that the predicted labels are not only accurate but as close to the true labels as possible. This method of classifying the instances into the nearest ordinal labels, is called ordinal regression (OR) [10,11], and the overall evaluation model is ordinal regression for doctor performance evaluation (OR-DPE).

In supervised learning OR resides between multi-classification and metric regression. The difference between the two is that the labels of the latter are in a limited but unordered set. The difference between OR and metric regression is that the OR labels do not represent numerical values. Although standard multi-classification and regression algorithms can be used to solve OR problems, they ignore the ordering information between labels. Some researchers [11,12] have proved that ordering information benefits modelling greatly. There exist many models especially designed for OR. The "Proportional Odds" model, designed in 1980 [13], is one of the earliest such models. Since then a wide range of OR models have been proposed including support vector machine (SVM)-based models [10,14-17], Neural Network-based models [18], Gaussian Process models [19], and more. An excellent survey [11], provides a comprehensive literature review about OR. The SVM-based model is one of the most popular models used in the field.

In the OR-DPE model, the consultation text is the input, and one label from the set of (1) very unsatisfied, (2) unsatisfied, (3) neutral, (4) satisfied, and (5) very satisfied is the output. The workflow of the OR-DPE model is shown in Figure 1. OR-DPE comprises of text preprocessing, representation, model training, and predictability. Because the communication between doctor and patient is through a text message, the DPE task is like text mining. The consultation texts are preprocessed and displayed as high dimensional vectors. Because the SVM-based model with linear kernel [14] performs excellently on large-scale data and is well suited for text mining fields, this model is preferred to address the DPE. In this paper, a new SVM-based Ordinal Partitioning model (SVMOP) is proposed as the OR model for DPE. With the SVMOP model, OR-DPE can, not only make sure that the predicted labels are as correct as possible, but also ensure that the incorrect labels are as close to true as possible. To our knowledge, this is the first time that the issue of DPE has been conceptualized as an ordinal regression task. Empirical studies on real data sets from one of the largest mobile doctor/patient communication platforms in China show that the model can achieve state-of-the-art performance from multiple metrics.

Figure 1. The general workflow of the ordinal regression for doctor performance evaluation (OR-DPE) model.



Methods

Preprocessing and Text Representation

The original corpus should be preprocessed, and each sample should be represented as an input vector. In the preprocessing step, punctuation and stop words will be removed. If the experimental data is written in Chinese, the words must be segmented as in Chinese text. Sentences are represented as character strings without natural delimiters. Chinese Word Segmentation (CWS) is used to identify word sequences in a sentence and mark boundaries in appropriate places. For example, CWS can put the character sequence "天花" together as a Chinese word for "smallpox" rather than the individual Chinese character "天" (sky) and "花" (flower) respectively. Word segmentation is a preliminary and important step for preprocessing. Most methods take the CWS as a sequence labeling problem [20], which can be formalized as supervised learning methods with customized features. Additionally, domain dictionaries with technical terms as ancillary resources, are beneficial for CWS and medical feature extraction. Here, 3 medical dictionaries are employed; one for Illness, one for Symptoms and one for Medicine. Most terms in the dictionaries are customized by medical experts and extended with new word detection techniques. We have collected 49,758 illness and symptom terms and 24,975 medical terms. Information about the dictionaries are shown in Table 1. For this purpose, we combined the dictionaries with Jieba tool, an open sourced Chinese segmentation software, for word segmentation.

For text representation, each sample is represented as an input vector where each dimension of the vector represents a feature. The element is the corresponding feature value. Feature engineering plays an important role in text mining. Apart from the basic text features such as Bag of Words (BOW) [21], unigrams, and bigrams, the custom medical features that can mirror some characteristics of the platform are utilized. These are specifically designed for the doctor/patient communication platform by domain experts and most are based on medical dictionaries. Typical text and medical features used in OR-DPE are presented in Table 2. Customized features (F1-F8) can capture domain knowledge: the count of medicine and symptom names in doctors' answers reflects the doctors' professional level; the number of Chinese characters in doctors' answers mirrors the service attitudes, and more. Likewise, the text features (F9 and F10) cover most consultation information. The feature value is the numerical value of the feature while the feature value of text features is the term frequency inverted document frequency (TF-IDF) [22]. TF-IDF reflects how important a word is to a document. If a word occurs rarely but appears frequently in a sample, it is most likely to reflect the characteristics of this sample. Specifically, TF-IDF is the product of two statistics: term frequency and inverse document frequency, where the former represents the frequency and the latter represents the inverse frequency of occurrence in all samples.

Table 1. The details about the medical dictionaries. "1≤terms≤3" means the number of terms having a character length less than 3 but greater than 1.

Number of phrases	Dictionary Name	
	Illness and Symptom Dictionary (N=49,758)	Medicine Dictionary (N=24,975)
1≤terms≤3, n (%)	32840 (66.00)	3746 (15.00)
4≤terms≤6, n (%)	16918 (34.00)	14486 (58.00)
terms≥7, n (%)	0 (0)	6743 (27.00)
Representative examples	神经衰弱症 (Neurosis), 高血压 (HTN), 天花 (smallpox)	帕罗西汀 (Paroxetine), 盐酸环苯扎林 (Flexeril)

Table 2. F1-F8 represent the customized medical features, while F9 and F10 are the text features.

Feature	Description
F1	The number of symptom names in doctors' answers
F2	The number of illness names in doctors' answers
F3	The number of medicine names in doctors' answers
F4	The number of patients' questions
F5	The number of doctors' answers
F6	The response time for the patient's first question
F7	The number of Chinese characters in patients' questions
F8	The number of Chinese characters in doctors' answers
F9	Unigrams
F10	Bigrams

The quantity of text features is so large that the customized features (see Table 2) can easily be overshadowed. To highlight the importance of customized features, they are boosted by the Gradient Boosting Decision Tree (GBDT) [23]. GBDT is a powerful tool in many industrial communities [24]. GBDT mines the most effective features and feature combinations by a decision tree to boost the performance of regression and classification tasks. This technique is applied to increase the number of custom medical feature combinations. The main idea of GBDT is to combine weak learners into a single, strong learner like other boosting methods. GBDT is an iteration algorithm, which is composed of multiple decision trees. In the *m*-th iteration of GBDT, assumes that there are some imperfect models, F_m. The GBDT would construct a better model F_{m+1} to approach the best model by adding an estimator *h*, namely $F_{m+1} = F_m(x) + h(x)$. Then the problem is transformed by the question of how to find $h(x)$. As the above equations imply, a perfect *h* should satisfy the equation:

$$h(x) = F_{m+1} - F_m(x) \approx y - F_m(x)$$

where *y* is the true label, $y - F_m(x)$ is called a loss function. In practice, a general way is to apply square loss function is: $\frac{1}{2}(y - F_m(x))^2$. Because the residual is exactly the negative gradients of the squared loss function. The problem on the left can then be solved directly by gradient descent algorithms. In our work, we apply GBDT to boost the 8 customized features shown in Table 2 to generate several effective feature combinations. According to the statistics, the number of features is 363,336 with text features, and 363,344 if adding the 8 customized features. After boosting the customized features, the number becomes 370,858. Another 7514 combined customized feature combinations have been added. The performances of various features are shown in Section Results.

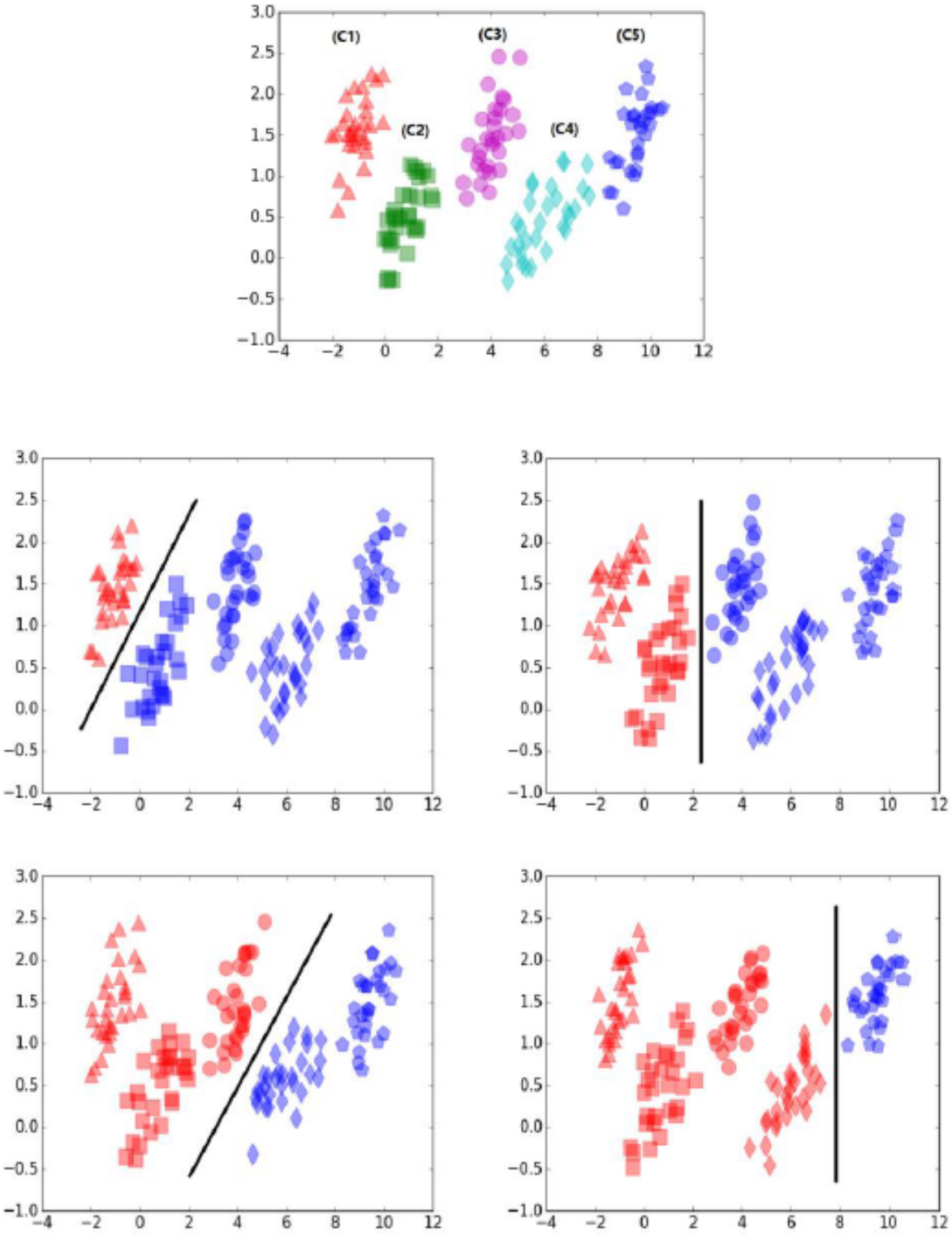
Model Training

How the Ordinal Regression Method for the Ordinal Regression for Doctor Performance Evaluation Model Was Chosen

There are many different models of OR. Referring to an OR survey [11], the models are grouped into three categories, namely the (1) naive approach, (2) threshold approach, and (3) ordinal partitioning approach. These models have corresponding strengths and weakness. The naive approach considers OR naively, as a standard classification task or a regression task [14,25]. At the same time, the ordering information between labels has been ignored. The threshold approach is based on the idea of approximating a real value predictor and then dividing the real line into intervals [10,15,26,27]. Assuming *P* is the number of categories, the objective of threshold-based OR models is to seek *P* –1 parallel hyperplanes further dividing the data into ordered classes. The ordinal partitioning approach uses the ordering information to decompose the ordinal regression into several binary classification tasks. For binary classification, there are many models to choose from. For example, Frank and Hall [16], applied decision trees as submodels while Waegeman and Boullart [17] used weighted SVMs as binary classifiers.

Since the ordering of information is conducive to model building [11], we chose the OR model from the latter two methods. As the number of samples is large and the dimension of the representative vectors is high, a model was chosen that can handle large-scale and high dimensional data. So, the ordinal partitioning approach is used instead of the threshold approach for OR problems depending on paralleled hyperplanes. There are many binary classifiers that can be chosen from the submodels. Hsieh et al [14] showed that the linear SVM is a robust tool that can deal with large-scale and high dimensional data. Inspired by these, we want to combine SVM with Ordinal Partitioning (SVMOP) as the OR model for the OR-DPE.

Figure 2. The demo that shows how a combined support vector machine and ordinal partitioning scheme model (SVMOP) works on ordinal data.



SVMOP Model and Training Algorithm

The OR problem can be described as follows: given a training set where x R^l is the i-th input vector (i=1,2,…,n), where n is the number of instances, l is the number of features, and y_i Y_i is the label of x_i. Assuming there are P categories and without loss of generality, we take the label set Y={1,2,…P}. The goal of OR is to find a function f: $X \rightarrow Y$ to predict the label of a new instance x. As mentioned earlier, SVMOP will be embedded into the OR-DPE model. Figure 2 illustrates the SVMOP procedure. In this figure, five ordinal categories of data are represented by different colors and shapes. The idea of SVMOP is to partition the overall model into P –1 binary classifications. Then the associated question: "Is the rank of the input greater than p?" can be asked. Here p=1,2,…,P –1. Therefore, the rank of x can be determined by a sequence of these binary classification problems. Specifically, when training the p-th binary classifier, the label y_i is retransformed to a new class

label depending on whether the label $\hat{y}_{pi}$ is greater than p or not, namely:

where i=1,2,…,n. Therefore, the problem can be reformulated: given a training set  , where x R^l is the i-th input sample, $\hat{y}_{pi}$ {-1,1} is defined by equation 1. The model aims to find a function to predict the ordered labels of new instances.

Linear SVM is one of the best candidates among the binary classifiers dealing with high dimensional data. Then linear SVM is taken as the p-th sub-model:

Where w_p represents the parameter of the p-th submodel, ξ_{pi} is the slack variable of the p-th submodel. As for the optimization solver, we chose the Dual Coordinate Descent algorithm (DCD) as the training algorithm of SVM [14]. DCD is one of the most effective training algorithms for linear SVMs. It solves the model in equation 2 by the Lagrange dual form. The dual form of the p-th sub-model in equation 2 is given as equation 3. Without loss of generality, we ignore the subscript p in the dual form:

where  is to employ a classic divide-and-conquer method for optimizing high dimensional problems. It starts from an initial zero vector α^0=**0** and generates a sequence of vectors  . For each iteration step, the algorithm sequentially selects one dimension associated with α to optimize by fixing other dimensions. Suppose α^* is the solution of equation 3 then the optimal value of w_p for equation 2 can be computed as follows:

Model Prediction

For model prediction, the research [11] shows that it is important to construct an effective rule for predicting new instances in the ordinal partitioning-based OR models. Many existing ways are based on the probability manipulation or outcomes by submodels to predict the label of a new instance. In the work by Frank et al [16], when estimating the probabilities for the first and the last class, the authors were dependent on a corresponding classifier. However, it needs to rely on two adjacent classifiers when computing the middle classes. This prediction method is simple and easy to implement, but may lead to a negative probability [11,28]. Another example in the work [17], the authors combined the outcomes of all the submodels to predict the label of a new instance x. However, their prediction function may cause ambiguities for some test samples.

To alleviate the problem with the above prediction functions, we propose a new prediction function as shown in equation 5:

where $r(x)$=1 if none of $w^T_p x$ is greater than 0. This prediction function relies on the discriminant planes and joins all binary classifiers to obtain a single classification. The p-th binary classifier provides the answer to the associated question: "Is the rank of the input x greater than p?", where p=1,2,…,P –1. That is, for prediction, the new sample x would be asked by a sequence of the questions above. And last, the predicted label equals $r(x)$ which represents the satisfaction degree. The greater $r(x)$, the more satisfied.

Statistical Methods and Evaluation Metrics

To better highlight the characteristics of ordinal regression models, we evaluated the performance with the following three common evaluation measures: (1) mean absolute error (MAE) [10,11,29], (2) mean square error (MSE) [30,31,32], and (3) pairwise accuracy (PAcc) [29,33,34]. MAE and MSE can directly measure the degree of deviation between the true label ($gold_i$) and predicted label (*predicted*). They can be defined by the following equations:

Since they are metrics measuring the error, the lower they are, the better their performance. PAcc is widely applied in the medical data analysis, ranking and statistics fields with the name of concordance index or Kendall τ [34,35]. PAcc could reflect the correct ratio of ranking between pairwise instances. Specifically, the set of preference evaluation pairs is represented as S,S={(i,j) | $gold_i > gold_j$ }.

The PAcc is given by

where "| S |" represents the number of the set S. It accords with the rule: the greater, the better.

Mining Predictive Features

Apart from rating doctors' performance, we continue to explore the most predictive features among text features and customized features in DPE. In general, predictive features always play significant and instructive roles on the platform construction. In this case, the most important features were extracted by analyzing the weight matrix W $R^{l\,x\,(P-1))}$, where l and P –1 are the dimensions of the matrix. As mentioned, l is the total number of all the features (that is, l=363, 344) and P is the number of categories, where P –1 is the number of the submodels. In equation 2, W is composed of the weight parameters, with w in each submodel, namely W=(w_1,w_2,…,w_{P-1}). We denote $W(j,:)$ as the j-th row vector and the absolute value of the elements in the row vector represents the contributions to each submodel for the j-th feature. The larger the value is, the more predictive property the feature has. For every feature in each kind of text feature or customized feature, described in Table 2, it owns its corresponding weight vector $W(j,:)$, where $1 \le j \le l$. We compute the total contribution Con_j of the j-th feature to the model decision by equation 9:

where "// //$_2$" represents the L2-norm of a vector. When the contributions of all the features have been computed, they would be ranked and hence obtain the top-most predictive features.

Results

Preparation of Datasets

To validate the proposed model on real data, the data from one of the popular doctor/patient consultation platforms (Platform X) in China was chosen as the experimental data sets. In practice, the platform maintains long-term cooperation with us. However, in order to comply with the confidentiality agreement with the platform, we are not able to use the real name of the platform in the paper but instead we use the name Platform X throughout. On Platform X, the consultation mainly consists of patients' questions (eg, "医生您好，如何补钙" [Hi doctor, how can I add more calcium?]) and the response (eg, "很高兴为您服务，您可以通过牛奶，豆制品，鱼等食物，也可以口服碳酸钙和葡萄酸钙" ["Glad to help you," "you can eat foods such as milk, bean products, and fish" and "You can also take calcium gluconate and calcium carbonate directly"]). To introduce experimental data, an actual consultation letter was used (Figure 3). Based on analysis of patients' questions, multiple question types are proposed. Most questions are about ailments that are not serious or related to personal privacy, like chronic pharyngitis and dermatosis. And, because they are flexible and convenient, most consultations are done through mobile software applications. Platform X faces the same problem when evaluating doctor's online performance. Platform X did not receive direct customer ratings or feedback since most patients tend to rate the very good or bad and at times feedback was not received because, for example, a customer may have been offline.

Of a sampling of 2,337,828 clinical data collected over the last twenty days, only 841,618 (36%) of the data was labeled by patients, which proves that most patients do not like to provide feedback. From the labeled data, only 720 instances, 1712 instances, and 8737 instances were labeled very unsatisfied, unsatisfied, and neutral respectively. The unbalanced data challenges the model. To alleviate the issue of unbalanced data and collect more instances of "very unsatisfied," we chose sample training data from Platform X's database. As previously mentioned, we have a long-term association with Platform X. It takes approximately two hours to access the entire database. Data collected is valid for about 18 months. After that, the same number of instances from each category are sampled. After filtering the data, (removing consultations with a length of less than 10 words), we have approximately 112,485 instances to use as experimental datasets. Each category contains approximately 22,497 samples which are randomly split into five sections with four sections serving as the training sets and the remaining one as the test set.

Baseline Methods

To better reflect the effectiveness of the proposed model with the above metrics, the following baseline methods will be compared with our model. These methods are popular and representative in OR fields. To solve the high dimensional data efficiently, the following models all use the linear kernel. The DCD algorithm is adopted to solve the following models. These are implemented by modifying the open source package LIBLINEAR [36] directly and all the codes related to the experiment are uploaded to a Github website. The following methods were used to compare with our model.

1. SVC [14]: Support vector classification with one versus the rest. This model belongs to the naive approach.
2. SVR [37]: Support vector regression. The ordinal labels are treated as continuous values. When predicting new instances, the predictions for test instances are rounded to the nearest ordinal label and the model belongs to the naive approach.
3. LR [38]: Logistic regression one versus the rest. This model belongs to the naive approach.
4. SVOR [10]: Support vector ordinal regression. This model aims to optimize multiple thresholds to define parallel discriminant hyperplanes. The SVOR model is used with implicit constraints and belongs to the threshold approach.
5. RedSVM [39]: Reduction support vector machine. A threshold approach and it is a reduction framework from ordinal ranking to weighted binary classification by extending examples.

Evaluation Performance

First, we compare the performance of five different baselines with our SVMOP model using different sets of features, including (1) text features (T), (2) text and customized features (T+C), and (3) text, customized, and boosted features (T+C+B). Three metrics, (ie, MAE, MSE, and PAcc) are used to evaluate model performance. Table 3 shows the results of the experiment. The best performance for each metric is represented by the footnote "k" while the best "one of" feature sets is represented by the footnote "e." In Table 3, the SVMOP model outperforms other baselines on MAE, MSE and PAcc with each type of feature sets, demonstrating the effectiveness of our model. On different set of features, all models achieved better performance with feature set T+C and feature set T+C+B. Furthermore, compared with feature set T+C+B, feature set T+C attained more improvement. In other words, using customized features are important for performance improvement

Figure 3. An example of consultation letters on Platform X. The left subfigure is the real consultation on Platform X by mobile software applications but without sensitive information such as doctors' photos. The right one is the version in English.

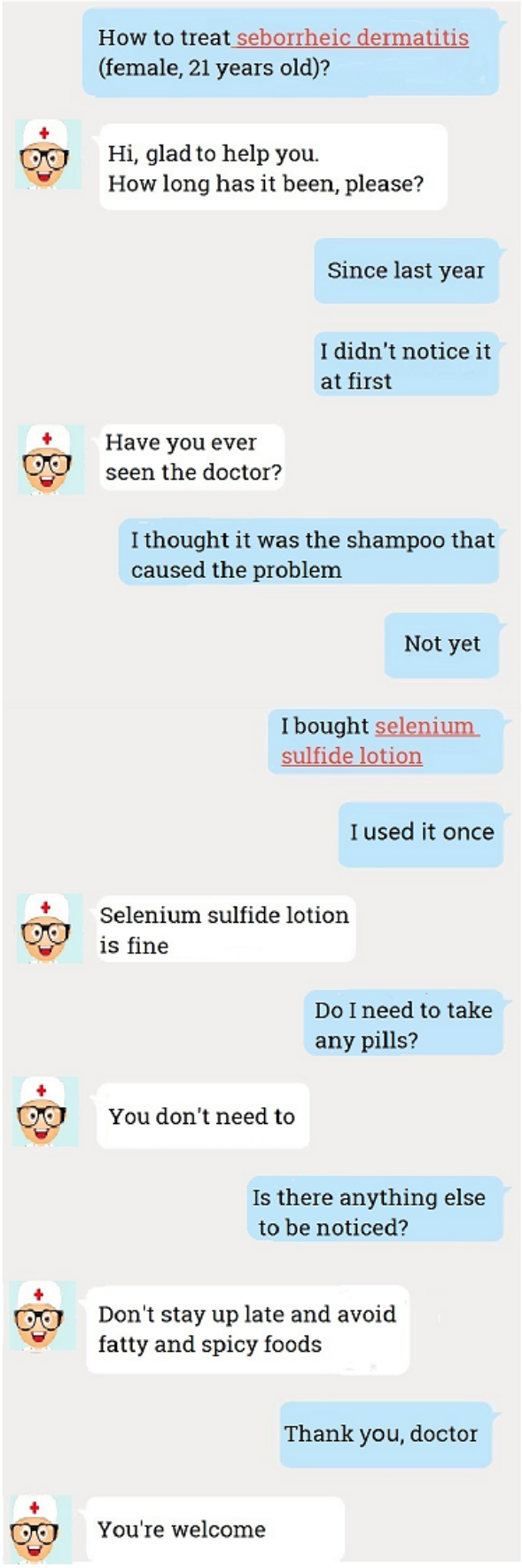


Table 3. Performances of various models having multiple feature sets (T, T+C, T+C+B) are shown in this table.

Method	Text (T)			Text and Customized (T+C)			Text, Customized, and Booster (T+C+B)		
	MAE[a]	MSE[b]	PAcc[c] (%)	MAE	MSE	PAcc (%)	MAE	MSE	PAcc (%)
SVC[d]	0.7925	1.7613	53.32	0.6726	1.3759	57.32	0.6212[e]	1.1981[e]	59.05[e]
SVR[f]	0.8023	1.3302	49.74	0.7050	1.1106	54.24	0.6906[e]	1.0332[e]	56.37[e]
LR[g]	0.7716	1.6883	53.86	0.6359	1.2606	57.77	0.5978[e]	1.1310[e]	59.50[e]
SVOR[h]	0.8086	1.3742	49.58	0.7170	1.1167	54.09	0.6665[e]	1.0143[e]	57.20[e]
RedSVM[i]	0.8046	1.3715	50.11	0.7168	1.1127	54.00	0.6718[e]	1.0236[e]	57.21[e]
SVMOP[j]	0.7054[k]	1.2706[k]	54.11[k]	0.6130[k]	1.0108[k]	57.92[k]	0.5864[e,k]	0.9605[e,k]	59.65[e,k]

[a]MAE: mean absolute error.

[b]MSE: mean standard error.

[c]PAcc: pairwise accuracy.

[d]SVC: support vector classification.

[e]Best "one of" feature sets.

[f]SVR: support vector regression.

[g]LR: logistic regression.

[h]SVOR: support vector ordinal regression.

[i]RedSVM: reduction support vector machine.

[j]SVMOP: a combined support vector machine and ordinal partitioning scheme model.

[k]Best performance for each metric.

Figure 4. Mean absolute error (MAE), mean square error (MSE), and pairwise accuracy (PAcc) varying from different models and different feature sets. LR: logistic regression; RedSVM: reduction support vector machine; SVC: support vector classification; SVMOP: a combined support vector machine and ordinal partitioning scheme model; SVOR: support vector ordinal regression; T: text features; T+C: text and customized features; T+C+B: text, customized, and boosted features.

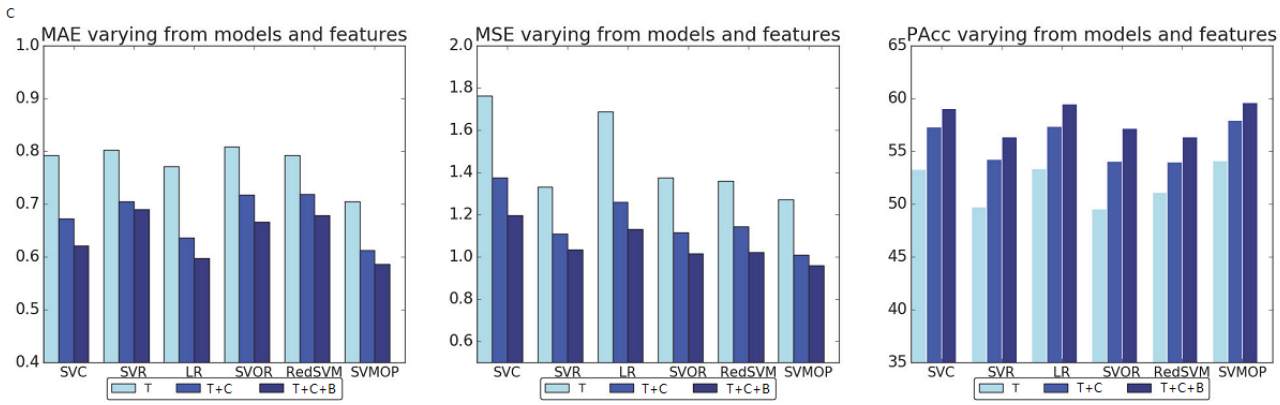


Figure 4 displays the performances of 6 models on 3 measures, namely MAS, MSE, and PAcc. As we can see, SVMOP greatly outperforms the other models on MAE, MSE, and PAcc. Additionally, the models that consider ordering information, namely, SVOR, RedSVM, and SVMOP, perform better than the rest on MSE; and SVC and LR achieve comparable performances with SVMOP on PAcc. To investigate the influence of the parameter, we show the various performance of each model as we change the parameter $\log_2 C$ in a range [-5,5]. In Figure 5, we find that the performances vary as the parameters change and the model can achieve the best performance in this range.

Additionally, the confusion matrices were used to further discuss the differences among the performance of different models. Each confusion matrix is generated by the corresponding model on feature set T+C+B. As shown in Figure 6, models that consider ordering information, such as SVOR, RedSVM, and SVMOP, misclassify the incorrectly labeled samples into the closest categories. For example, in the confusion matrix of SVMOP, the third cell of the third row shows that 63% of most (17%) misclassified instances fell into Category 2. In contrast to the confusion matrix of SVMOP, when looking at the third row of the confusion matrix of SVC, we find that most (19%) of the misclassified instances fall into Category 1. For this study, this example illustrates that the nonordering information methods, such as SVC and LR, can misclassify doctors having *neutral* performance levels into the *very unsatisfied* category. However, methods that consider ordering information of doctors' performances, such as SVOR, RedSVM and SVMOP,

are more likely to place misclassified neutral doctors into the *unsatisfied* category.

Predictive Features Analysis

As for mining the most predictive features, (see equation 9), and after computing and sorting each feature, we find that 4 customized features are in the top 5 most predictive features, including F3, F4, F5, and F7. Feature F6 ranks 8th and feature F8 ranks 23rd. In other words, most of the customized features play the most predictive roles in DPE, which is consistent with our numerical results. In view of the most predictive text features, we find that the features that contribute most to the model decision are the polite expressions like: “不客气” (“You’re welcome”), “谢谢您” (“Thank you”), “很乐意帮助您” (“Glad to help you”) and sensitive words such as “好评” (“good rating”), “态度”(“attitude”), “五星”(“five-star”). Some words like “禁忌辛辣食物” (“avoid spicy foods”), “对身体有害” (“bad for health”) are helpful, and indicates that the doctor is explaining some issues in more detail. These features cannot guide in questionnaire design but are beneficial for platform building and optimization.

Figure 5. The different performances with different parameters in training process with the text, customized, and boosted feature set (T+C+B). LR: logistic regression; MAE: mean absolute error; MSE: mean square error; PAcc: pairwise accuracy; RedSVM: reduction support vector machine; SVC: support vector classification; SVMOP: a combined support vector machine and ordinal partitioning scheme model; SVOR: support vector ordinal regression; SVR: support vector regression.

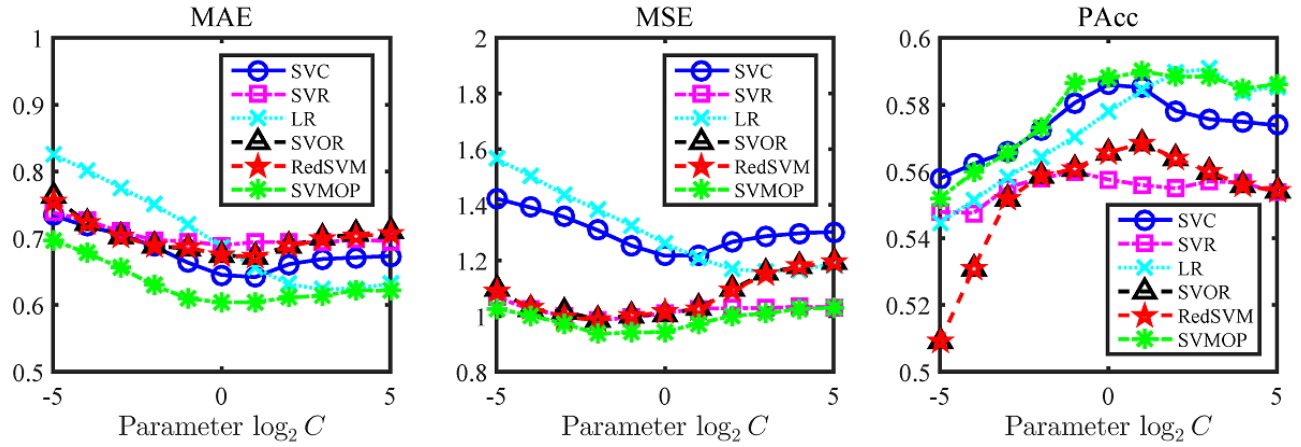
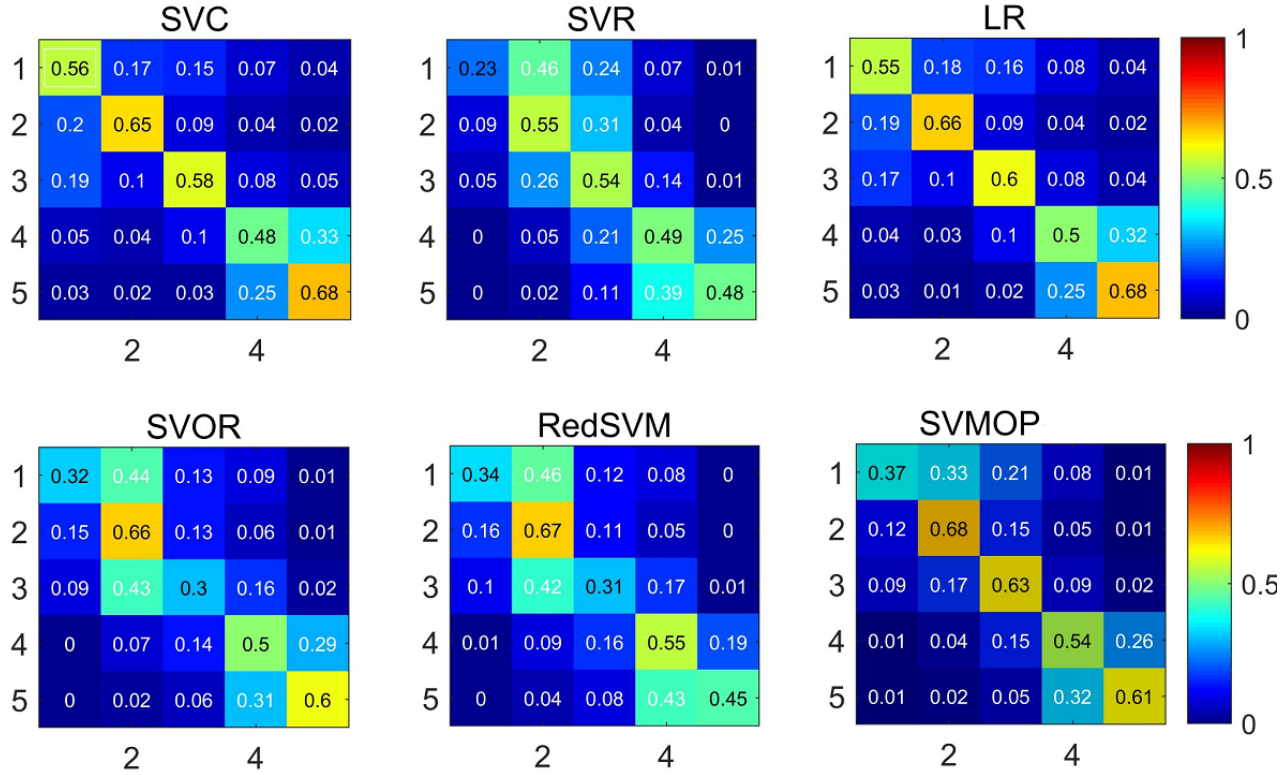


Figure 6. The confusion matrices of different models with the text, customized, and boosted feature set (T+C+B). RedSVM: reduction support vector machine; SVC: support vector classification; SVMOP: a combined support vector machine and ordinal partitioning scheme model; SVOR: support vector ordinal regression; SVR: support vector regression.



Discussion

Principal Results

Statistically, 36% of data on doctor/patient communication platforms has been labeled by patients, showing that 64% of the clinical data lack evaluation. Considering that doctors' performance could affect patient satisfaction, we take the DPE task as an ordinal regression problem, ensuring the automatically predicted labels are as close as possible to the true ones. The OR-DPE, SVMOP model with revised prediction is applied as the core model, and the metrics of MAE, MSE, PAcc, and SVMOP models with feature set T+C+B could achieve state-of-the-art performance. Compared with other auto-evaluation models, SVMOP improves MAE by 0.1, MSE by 0.5, and PAcc by 5%. The customized features improve MAE by 0.1, MSE by 0.2, and PAcc by 3%. Additionally, with the boosting technique, the performance of SVMOP can be further improved. Furthermore, based on OR-DPE model, predictive features like polite expressions and sentiment words can also be mined, which can be used to guide the development of mHealth platforms.

Comparison with Prior Work

The experiments conducted on real data have validated the effectiveness of SVMOP. Because of the noise in the real data, we continue to experiment on benchmark OR datasets [19] in a precisely controlled environment. The datasets can be downloaded from the public website. In this experiment, we compare our model with all the baselines mentioned in the paper. The details about the benchmark datasets and the results are shown in Tables 4 and 5.

We find that SVMOP always performs better than other baselines on MAE, MSE, and PAcc. The results verify the effectiveness of SVMOP on clean data. Therefore, the good results benefit from the SVMOP model but not the experimental data about DPE, which further demonstrates the correctness of choosing SVMOP as the core model of OR-DPE.

Limitations

Although this study has solved the problem of doctors' auto-evaluation on doctor/patient communication platforms by the ordinal regression approach, there are limitations. Firstly, the definition of a good consult here is related to user satisfaction, not to medical accuracy or clinical utility. A good doctor seems to be a likable one, but a likable one may make incorrect medical decisions. Secondly, Farmer et al [40] point out that doctors' work should be evaluated by multiple complex professional factors. In other words, a good consult is not only related to patients but also to many other factors. One way to handle this issue is to multisource feedback [41], which is called 360-degree evaluation in which key performance behaviors are simultaneously rated by peers, patients, and coworkers. Considering the characteristics of doctor/patient communication platforms, peer evaluation can be achieved by questionnaires, and the predictive features generated by the OR-DPE model may, in turn guide the questionnaire design.

Conclusions

The authors are the first to conceptualize the problem of DPE as an ordinal regression task and develop an OR-DPE model to address it. Apart from the basic text features, we use eight customized features suggested by domain experts as important features to improve model performance. Furthermore, we applied GBDT to boost the 8 customized features. Additionally, we proposed a new model called SVMOP which has a reasonable and effective prediction function. Experiments show that the performance of SVMOP is better than many other machine learning methods on MAE, MSE, and PAcc. In summary, with the OR-DPE model, the mHealth platform could not only make an auto-evaluation of online doctors' performance but also mine the most effective features which can then be further applied to guide the promotion of doctors and platforms. In the future, we hope our model can also be explored and applied to other medical service-oriented issues in medical education.

Table 4. Benchmark datasets. "#ins" is the number of instances. "#fea" is the number of features. "#class" is the number of classes.

Datasets	#ins	#fea	#class
housing-5	10120	14	5
machine-5	4180	7	5
abalone-5	83540	11	5
housing-10	10120	14	10
machine-10	4180	7	10
abalone-10	83540	11	10

Table 5. The mean absolute error (MAE), mean standard error (MSE), and pairwise accuracy (PAcc) performances of different models on benchmark datasets. The best result is indicated by a footnote.

Datasets	SVC[a]	SVR[b]	LR[c]	SVOR[d]	RedSVM[e]	SVMOP[f]
Mean absolute error (MAE)						
housing-5	0.517	0.454	0.435	0.398	0.403	0.366[g]
machine-5	0.606	0.550	0.451	0.390	0.424	0.369[g]
abalone-5	0.798	0.712	0.700	0.683	0.675	0.648[g]
housing-10	1.513	0.962	0.999	0.859	0.848	0.757[g]
machine-10	1.425	1.151	0.986	0.935	0.927	0.841[g]
abalone-10	1.959	1.451	1.557	1.435	1.434	1.391[g]
Mean standard error (MSE)						
housing-5	0.665	0.545	0.612	0.494	0.524	0.446[g]
machine-5	0.994	0.634	0.648	0.469	0.505	0.429[g]
abalone-5	1.450	0.992	1.244	1.042	0.991	0.962[g]
Housing-10	4.564	1.858	2.560	1.694	1.642	1.453[g]
machine-10	3.998	2.487	2.277	1.786	1.720	1.547[g]
abalone-10	7.222	3.703	5.091	3.586	3.783	3.635[g]
Pairwise accuracy (PAcc)						
housing-5	0.614	0.638	0.658	0.663	0.659	0.676[g]
machine-5	0.602	0.604	0.652	0.666	0.655	0.680[g]
abalone-5	0.547	0.553	0.584	0.584	0.577	0.589[g]
Housing-10	0.552	0.623	0.609	0.635	0.637	0.642[g]
machine-10	0.488	0.562	0.597	0.601	0.599	0.612[g]
abalone-10	0.514	0.568	0.566	0.565	0.568	0.569[g]

[a]SVC: support vector classification.

[b]SVR: support vector regression.

[c]LR: logistic regression.

[d]SVOR: support vector ordinal regression.

[e]RedSVM: reduction support vector machine.

[f]SVMOP: a combined support vector machine and ordinal partitioning scheme model.

[g]Best result.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 91546201, No. 71331005, No. 71110107026, No. 11671379, No. 11331012, No. 71501044), the Fundamental Research Funds for the Central Universities in UIBE (No. 16YQ07), UCAS Grant (No. Y55202LY00).

Conflicts of Interest

None declared.

References

1. Eysenbach G. What is e-health? Journal of Medical Internet Research 2001 Jun;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]

2. Borrelli B, Ritterband LM. Special issue on eHealth and mHealth: Challenges and future directions for assessment, treatment, and dissemination. Health Psychology 2015 Dec;34S:1205. [doi: 10.1037/hea0000323] [Medline: 26651461]
3. Jonathan G. 30 Amazing Mobile Health Technology Statistics for Today's Physician. URL: https://getreferralmd.com/2015/08/mobile-healthcare-technology-statistics/ [accessed 2017-10-01] [WebCite Cache ID 70SvHWknK]
4. Caruana A. Service loyalty: The effects of service quality and the mediating role of customer satisfaction. European journal of marketing 2002;36(7/8):811-828 [FREE Full text] [doi: 10.1108/03090560210430818] [Medline: 10786546]
5. Gremler DD, Brown SW. Service loyalty: its nature, importance, and implications. Advancing service quality: A global perspective 1996;5:171-181. [doi: 10.1017/CBO9781107415324.004]
6. Mosahab R, Mahamad O, Ramayah T. Service quality, customer satisfaction and loyalty: A test of mediation. International business research 2010;3(4):72. [doi: 10.5539/ibr.v3n4p72]
7. Overeem K, Faber MJ, Arah OA. Doctor performance assessment in daily practise: does it help doctors or not? A systematic review. Medical education 2007 Dec;41(11):1039-1049. [doi: 10.1016/j.ijmedinf.2006.09.001] [Medline: 17049917]
8. Li P, Yan Y, Wang C, et A. Customer voice sensor: A comprehensive opinion mining system for call center conversation. 2016 Presented at: IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA); 2016; Chengdu, China p. 324-329. [doi: 10.1109/ICCCBDA.2016.7529578]
9. Paprzycki M, Abraham A, Guo R, et A. Data mining approach for analyzing call center performance. Innovations in Applied Artificial Intelligence 2004:1092-1101. [doi: 10.1007/978-3-540-24677-0_112]
10. Chu W, Keerthi SS. Support vector ordinal regression. Neural computation 2007;19(3):792-815 [FREE Full text]
11. Gutiérrez PA, Pérez-Ortiz M, Sanchez-Monedero J. Ordinal regression method survey and experimental study. IEEE Transactions on Knowledge and Data Engineering 2016;28(1):127-146 [FREE Full text]
12. Huhn JC, Hullermeier E. Is an ordinal class structure useful in classifier learning? International Journal of Data Mining, Modelling and Management 2008;1(1):45-67. [doi: 10.1504/IJDMMM.2008.022537]
13. McCullagh P. Regression models for ordinal data. Journal of the royal statistical society 1980;42(2):109-142 [FREE Full text]
14. Hsieh CJ, Chang KW, Lin CJ, et A. A dual coordinate descent method for large-scale linear SVM. 2008 Presented at: Proceedings of the 25th international conference on Machine learning. ACM; 2008; New York, USA p. 408-415 URL: https://doi.org/10.1145/1390156.1390208
15. Chu W, Keerthi SS. New approaches to support vector ordinal regression. 2005 Presented at: The 22nd international conference on Machine learning; 2005; New York, USA p. 145-152 URL: https://doi.org/10.1145/1102351.1102370
16. Frank E, Hall M. A simple approach to ordinal classification. 2001 Presented at: The European Conference on Machine Learning; 2001; United Kingdom p. 145-156. [doi: 10.1007/3-540-44795-4_13]
17. Waegeman W, Boullart L. An ensemble of weighted support vector machines for ordinal regression. International Journal of Computer Systems Science and Engineering 2009;3(1):47-51 [FREE Full text]
18. Deng WY, Zheng QH, Lian S, et A. Ordinal extreme learning machine. Neurocomputing 2010;74(1):447-456 [FREE Full text]
19. Chu W, Ghahramani Z. Gaussian processes for ordinal regression. Journal of machine learning research 2005;6(3):1019-1041.
20. Xue N. Chinese word segmentation as character tagging. International Journal of Computational Linguistics & Chinese Language Processing, Special Issue on Word Formation and Chinese Language Processing 2003;8(1):29-48 [FREE Full text]
21. Zhang Y, Jin R, Zhou Z. Understanding bag-of-words model: a statistical framework. International Journal of Machine Learning and Cybernetics 2010;1(1-4):43-52. [doi: 10.1007/s13042-010-0001-0]
22. Witten IH, Frank E, Hall MA, et A. Data Mining: Practical machine learning tools and techniques. In: Morgan Kaufmann, ELSEVIER. Cambridge: ELSEVIER; 2016.
23. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics 2001;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
24. Chapelle O, Shivaswamy P, Vadrevu S, et A. Multi-task learning for boosting with application to web search ranking. 2010 Presented at: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2010; Washington, DC, USA p. 1189-1198 URL: https://github.com/tqchen/xgboost [doi: 10.1145/1835804.1835953]
25. Kramer S, Widmer G, Pfahringer B, et A. Prediction of ordinal classes using regression trees. Fundamenta Informaticae 2001;47(1-2):1-13. [doi: 10.1007/3-540-39963-1_45]
26. Herbrich R, Graepel T, Obermayer K. Support vector learning for ordinal regression. 1999 Presented at: The Ninth International Conference on Artificial Neural Networks; 1999; Edinburgh, UK p. 97-102.
27. Gu B, Sheng VS, Tay KY, Romano W, Li S. Incremental Support Vector Learning for Ordinal Regression. IEEE Transactions on Neural networks and learning systems 2015;26(7):1403-1416. [doi: 10.1109/TNNLS.2014.2342533] [Medline: 25134094]
28. Cardoso JS, Costa JF. Learning to classify ordinal data: The data replication method. Journal of Machine Learning Research 2007;8(7):1393-1429 [FREE Full text]
29. Gutiérrez PA, Pérez-Ortiz M, Fernández-Navarro F. An experimental study of different ordinal regression methods and measures. 2012 Presented at: International Conference on Hybrid Artificial Intelligence Systems; 2012; Berlin, Heidelberg p. 296-307. [doi: 10.1007/978-3-642-28931-6_29]

30. Diao Q, Qiu M, Wu CY, et A. Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars). 2014 Presented at: The 20th ACM SIGKDD international conference on Knowledge discovery and data mining; 2014; New York URL: https://doi.org/10.1145/2623330.2623758
31. Shimada K, Endo T. Seeing several stars: A rating inference task for a document containing several evaluation criteria. Advances in Knowledge Discovery and Data Mining 2008:1006-1014. [doi: 10.1007/978-3-540-68125-0_106]
32. Tang D, Qin B, Liu T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. 2015 Presented at: International Conference on Empirical Methods in Natural Language Processing; 2015; Lisbon, Portugal p. 1422-1432. [doi: 10.18653/v1/D15-1167]
33. Kuo TM, Lee CP, Lin CJ. Large-scale kernel rankSVM. 2014 Presented at: Proceedings of the 2014 SIAM international conference on data mining; 2014; Philadelphia, Pennsylvania, USA p. 812-820. [doi: 10.1137/1.9781611973440.93]
34. Lee C, Lin C. Large-scale linear rankSVM. Neural computation 2014;26(4):781-817. [doi: 10.1162/NECO_a_00571] [Medline: 24479776]
35. KENDALL MG. A new measure of rank correlation. Biometrika 1938;30(1/2):81-93. [doi: 10.1093/biomet/30.1-2.81]
36. Fan RE, Chang KW, Hsieh CJ, et A. LIBLINEAR: A library for large linear classification. Journal of machine learning research 2008;9:1871-1874 [FREE Full text] [doi: 10.1145/1390681.1442794]
37. Ho CH, Lin CJ. Large-scale linear support vector regression. Journal of Machine Learning Research 2012;13:3323-3348 [FREE Full text]
38. Yu H, Huang F, Lin C. Dual coordinate descent methods for logistic regression and maximum entropy models. Machine Learning 2011;85(1-2):41-75. [doi: 10.1007/s10994-010-5221-8]
39. Lin H, Li L. Reduction from cost-sensitive ordinal ranking to weighted binary classification. Neural Computation 2012;24(5):1329-1367 [FREE Full text] [doi: 10.1162/NECO_a_00265] [Medline: 22295981]
40. Farmer EA, Beard JD, Dauphinee WD, LaDuca T, Mann KV. Assessing the performance of doctors in teams and systems. Medical education 2002;36(10):942-948 [FREE Full text] [doi: 10.1046/j.1365-2923.2002.01311.x]
41. Lockyer J. Multisource feedback in the assessment of physician competencies. Journal of Continuing Education in the Health Professions 2003;23(1):4-12 [FREE Full text] [doi: 10.1002/chp.1340230103] [Medline: 12739254]

Abbreviations

BOW: bag of words
CWS: Chinese word segmentation
DCD: Dual Coordinate Descent algorithm
DPE: doctor performance evaluation
GBDT: Gradient Boosting Decision Tree
LR: logistic regression
MAE: mean absolute error
mHealth: mobile health
MSE: mean square error
OR: ordinal regression
OR-DPE: ordinal regression for doctor performance evaluation
PAcc: pairwise accuracy
RedSVM: reduction support vector machine
SVC: support vector classification
SVM: support vector machine
SVMOP: a combined support vector machine and ordinal partitioning scheme model
SVOR: support vector ordinal regression
SVR: support vector regression
T: text features
T+C: text and customizable features
T+C+B: text, customizable, and boosted features
TF-IDF: term frequency inverted document frequency

Edited by G Eysenbach; submitted 30.10.17; peer-reviewed by S Allin, H Zhang, Y Liu; comments to author 15.03.18; revised version received 11.04.18; accepted 08.05.18; published 18.07.18.

Please cite as:
Shi Y, Li P, Yu X, Wang H, Niu L
Evaluating Doctor Performance: Ordinal Regression-Based Approach
J Med Internet Res 2018;20(7):e240
URL: http://www.jmir.org/2018/7/e240/
doi:10.2196/jmir.9300
PMID:30021708

Original Paper

Public Awareness, Usage, and Predictors for the Use of Doctor Rating Websites: Cross-Sectional Study in England

Salma Patel[1], BSc (Hons), PhD; Rebecca Cain[2], BA (Hons), PhD; Kevin Neailey[3], BSc (Hons), PhD; Lucy Hooberman[3], MA (Hons), FRSA

[1]School of Health and Society, University of Salford, Manchester, United Kingdom

[2]Loughborough Design School, Loughborough University, Loughborough, United Kingdom

[3]Warwick Manufacturing Group, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Salma Patel, BSc (Hons), PhD
School of Health and Society
University of Salford
The Crescent
Salford
Manchester, M5 4WT
United Kingdom
Phone: 44 (0)161 295 2394
Email: s.patel48@salford.ac.uk

Abstract

Background: With the advent and popularity of social media and consumer rating websites, as well as the emergence of the digitally engaged patient, there has been an increased interest in doctor rating websites or online patient feedback websites, both inside and outside academia. However, there is very little known about how the public across England views such rating websites as a mode to give patient experience feedback.

Objective: The aim of the overall study was to measure and understand public awareness, usage, and attitudes towards doctor rating websites as a mode to give experiential feedback about GPs in general practice in England. This paper reports on the findings of one of the aims of the study, which was to measure public awareness, current usage and future consideration of usage of online patient feedback websites, within the context of other feedback methods, This could allow the value of online patient feedback websites to be determined from the patients' perspective.

Methods: A mixed methods population questionnaire was designed, validated and implemented face-to-face using a cross-sectional design with a representative sample of the public (n=844) in England. The results of the questionnaire were analyzed using chi-square tests, binomial logistic regressions, and content analysis. The qualitative results will be reported elsewhere.

Results: Public awareness of online patient feedback websites as a channel to leave experiential feedback about GPs was found to be low at 15.2% (128/844). However, usage and future consideration to use online patient feedback websites were found to be extremely low, with current patient usage at just 0.4% (3/844), and patient intention to use online patient feedback in the future at 17.8% (150/844). Furthermore, only 4.0-5.0% of those who would consider leaving feedback about a GP in the future selected doctor rating websites as their most preferred method; more than half of patients said they would consider leaving feedback about GPs using another method, but not using an online patient feedback website.

Conclusions: The findings suggest that online patient feedback websites may not be an effective channel for collecting feedback on patient experience in general practice. Feedback on online patient feedback websites is not likely to be representative of the patient experience in the near future, challenging the use of online patient feedback not just as a mode for collecting patient experience data, but for patient choice and monitoring too. We recommend the National Health Service channels its investment and resources towards providing more direct and private feedback methods in general practice (such as opportunities for face-to-face feedback, email-based feedback, and web-based private feedback forms), as these are currently much more likely to be used by the majority of patients in England.

(J Med Internet Res 2018;20(7):e243) doi:10.2196/jmir.9523

KEYWORDS

online reviews; Physician quality; primary care; Internet; quality patient empowerment; quality transparency; public reporting

Introduction

Since the 1990s, there has been an exponential increase in the usage of the internet around the world, including a rise in the number of people using the internet for health purposes [1]. There has also been a growth in the number of people giving ratings and reviews online for products and services (such as on amazon.com). Some argue that this has allowed for transparent information and communication to influence change and has provided opportunities for consumers to read reviews and make more informed choices [2-4]

The National Health Service (NHS) when founded in 1948 was paternalistic in its approach to the care of patients [5]. However, from the 1970s onwards, there has been an increasing emphasis on patient and public involvement (PPI), with the introduction of multiple measures to collect patient experience feedback, and the provision of more patient choice [5-7]. There has also been a growing emphasis on public reporting of performance measures across the government, including healthcare. Patients are now argued to have an equal relationship with the NHS and other healthcare providers [5,8].

All of the above factors led to the evolution of online patient feedback (OPF) websites or doctor-rating websites. NHS England introduced an OPF website in 2007—the NHS Choices feedback website [9]. For primary care and general practice, this means that patients can use these types of websites to review their healthcare experience and use these reviews to choose a provider. The presence of these websites has been argued to yield multiple benefits, including empowering patients, improving transparency and enhancing patient choice [9-11]. However, there is little evidence to support these claims.

Despite this, there has been a growth in the volume of OPF, which may suggest that patients in England (and other parts of the world) are embracing the opportunity to review their health care online [11-14]. Similarly, growth in the development of OPF can also be seen, with the development of websites where patients can review their medication and treatment plan [15].

There has also been a steady increase in research into OPF websites, with studies conducted in the UK [9,11,13,14,16--24], Germany [16-21], Netherlands [22], Australia [23] and other countries [24,25] all contributing to the OPF evidence basis. Some evidence can be found to suggest that there is an association between online ratings and the quality of care [12,13,26-28], but the results are often conflicting [29].

Studies conducted outside of England have focused on the characteristics of patients that use OPF websites [18,30-33]. However, the findings cannot be directly applied to England due to the nature of the healthcare systems being distinctly different [10]. Furthermore, the main OPF website in England is a practice-based OPF website, where patients leave reviews under a practice name, rather than the name of the general practitioner (GP).

In England, 3 studies focused on OPF websites from the patients' perspective [34-36]. The first is a qualitative study based on 3 focus groups conducted by the Nuffield Trust which explored public attitudes towards health and social care ratings. The findings suggested that patients relied more on the word-of-mouth to choose a GP rather than an overall score of a GP [34].

The second was a small convenience survey study conducted with 200 participants in one borough of London [35] to explore the predictors for the usage of doctor-rating websites. The findings suggested a low awareness of doctor-rating websites. Those younger, or ethnically white, or those when deciding where to receive care either give importance to the reputation of the doctor or hospital statistics, are more likely to be aware of doctor-rating websites. They also found that income, ethnicity, and the doctor-patient relationship were significant predictors of future intention to use doctor-rating websites.

This latter study was small and was not representative of patients across England. More crucially, however, it was not evident from the study for which purpose patients were using or were aware of these websites (for feedback or choice or both). Furthermore, none of the studies found in the literature compared patient awareness, usage or predictors of OPF to other methods of collecting feedback that are available for patients to use. This means that it is difficult to truly determine usage or awareness outside of its context. Hence, for example, it may be that usage of other methods is also low too, and therefore limited usage is not exclusive to OPF websites. It is also not clear whether OPF is filling a feedback gap".

The authors of this paper, therefore, conducted a small qualitative study (n=18) to explore patients' views towards giving online feedback and ratings to GPs in England. This was done within the context of other feedback methods available in primary care, in particular, paper-based feedback cards, which has been published [36]. This current study is a follow up to that study [36], to explore nationwide public views towards online patient feedback or feedback on doctor-rating websites (both terms are used interchangeably in this paper) in England.

The aim of this study was to measure and understand public awareness, usage, and attitudes towards doctor-rating websites, within the context of other feedback methods. Understanding how patients perceive and use OPF websites in comparison to other feedback methods can help determine whether OPF websites are of any perceived value to patients. This may potentially even help increase usage of OPF websites and improve the design and user-experience of OPF websites. This also allows for adequate comparison and a more comprehensive understanding of public awareness and usage of doctor-rating websites, rather than an isolated one, as previous researchers in this field have conducted [18,31,32,35,37]. These researchers also explored the effect or association of socio-demographic variables and other health factors on the usage and awareness of doctor-rating websites and used some of the factors to explain the variation in results. This was also conducted in this study.

This study was also unique in that it focused specifically on using doctor-rating websites to give feedback about GPs, whereas all of the previous studies [18,31,32,35,37] explored doctor-rating websites more generally (for feedback and choice), and asked respondents to comment on its overall use for all healthcare services.

This paper addresses the research question: Are patients aware of OPF websites as a channel for experiential feedback in general practice, and do they use them? (The other mainly qualitative findings of this study will be reported elsewhere).

Methods

Questionnaire Design and Mode

A mixed methods population questionnaire was developed by the first author (SP) using the themes that emerged from the authors' qualitative study [36] and previous literature (see Multimedia Appendix 1 for a copy). It was evaluated and validated based on the Total Survey Error Framework [38] using 7 stages, which included multiple-stage expert reviews (n=16), cognitive interviews (n=9), and pilot testing (n=22). The study had ethical approval from the Biomedical and Scientific Research Ethics Committee at the University of Warwick (ref REGO-2015-1472; May 2015 and #REGO-2015-1472 AM01; Dec 2015).

A decision was made for Ipsos Mori (a research company) to implement the questionnaire face-to-face with a representative sample of the public across England. Face-to-face was the most appropriate mode because of the length of the questionnaire, it was within budget, and it is also least burdensome on the respondent [39]. Ipsos MORI was chosen because they are a reputable and well-experienced research company, who also conducts the national GP Patient Survey on behalf of NHS England (and the Department of Health).

Sample Size and Sampling Procedure

An target sample size of 850 members of the public (in England) was set based on guidance from Field [40] to allow prevalence statistical estimate proportions to be within 3.5% confidence interval with 95% confidence level. A post-hoc sample size analysis illustrated that the prevalence data was within a confidence interval of 3.37% with a 95% confidence level.

Random location quota sampling using quotas for age, working status, gender and tenure within the region were used in this study. There were 2 stages to the sampling. In the first stage of sampling, approximately 180 Local Area Authorities were randomly selected from all those in the UK, some of which were in Scotland and Wales and therefore do not feature in this study. In the second stage of sampling, one Output Area (a small area made up of around 60 to a 100 addresses) was randomly selected from each of the Local Area Authorities selected in the first stage. These were the output areas where interviewers went to conduct the interviews with the public. Interviewers (n=155) were given quotas of people to interview for each Output Area according to age, working status, gender and tenure within the region.

Data Collection Procedure

The questionnaire from this study (which was around 10 minutes long) was included in the Ipsos MORI Face-to-Face Omnibus survey called Capibus (which runs every week and is around 30 minutes long) and was conducted using the Computer Assisted Personal Interviewing technique (ie, face-to-face interviews assisted by a computer) by 155 trained interviewers in people's homes from January 29, 2016 to February 10, 2016. Informed consent was taken verbally from all respondents before entering their homes. Interviewers went door to door and invited the person who answered to take part. The visits were spread out during the week, including evenings and weekends.

During the interview, interviewers immediately noted down each response on to their laptops, and the results were collated in real-time and recorded centrally by Ipsos MORI. There were 110/844 (13%) of all interviews validated (back-checked) so that the interview data was validated according to the ISO 20252 guide

Data Preparation

The data captured was provided to the first author (SP) in an SPSS file and Excel files. There was no missing data because the computer programming of the script ensured all respondents answered the relevant questions.

Weighting the Quantitative Survey Data

The sample profile produced for this study was similar to that achieved on The National Readership Survey (NRS), which uses random probability sampling. Therefore, using rim-weighting, only a very small corrective weighting was applied (on gender, age, social grade, region, working status, tenure, and ethnicity) by Ipsos MORI to adjust the final results to make them in-line with the national demographic profile. This was so that any minor deficiencies or biases in the sample could be corrected and to ensure that the sample was as close to a nationally representative sample.

The unweighted and weighted profile data can be seen in Table 1, which shows minor differences between profiles. For the responses to the questions on the questionnaire, the overall responses between the weighted and unweighted data varied if at all by only 1% or 2%.

Data Analysis

IBM SPSS Statistics 22 was used to conduct the statistical analysis (content analysis was conducted on the qualitative data, and the results for which will be reported elsewhere). The sampling weights provided by Ipsos MORI were first applied to the data to correct for known sample biases. Univariate analysis or descriptive statistics was performed to describe respondent demographics, and responses to all other relevant questions.

Bivariate analysis was used to describe differences for the main variables (dependent variables, for example, awareness, usage) with the demographic characteristics (independent variables, for example, gender and age). All variables were categorical, and therefore a 2-tailed chi-square test (or Pearson's test where appropriate) was used, with <.05 considered to be statistically significant. The demographic independent variables (eg, gender

and age) were then included in binomial logistic regression models, which were adjusted manually to determine which demographic factors in combination had a signification association or were predictors for the dependent variable [41]. Results were presented as odds ratio and 95% confidence intervals, using the format recommended by Peacock and Kerry [42] for publication. The results for the first binomial logistic regression model and its interpretation were checked and approved by an experienced academic medical statistician in March 2016.

Results

Response Rate and Demographic Characteristics

A total of 844 respondents over the age of 15 years from England responded to the questionnaire. The sociodemographics that respondents were asked about included gender, age, social grade, region, qualification, income and ethnicity, and these are reported in Table 1, including both the weighted data used in the analysis as well as the unweighted data. There were 4 further questions related to internet usage and health also asked, and the responses to these are also listed in Table 1. These 11 demographic variables are the independent variables against which other dependent variables were checked for association during the analysis. Further details are in the forthcoming sections.

Results on Awareness

Awareness of the Opportunity to Give Feedback About Care From General Practitioners Using Any Method

A total of 326 of 844 (38.6%) respondents were aware that they could give feedback about their experience of receiving care from a GP, whereas 518 (61.4%) were not aware that they could give feedback at all.

The effect of 11 demographic variables (in Table 1) on awareness was explored using chi-square tests and binomial logistic regression. The following 4 variables (Textbox 1) remained significant (also see Table 2).

Awareness of Doctor-Rating Websites for Giving Feedback About Experience of Receiving Care From General Practitioners

All respondents were provided an explanation of doctor-rating websites on screen and verbally by the interviewer (see Multimedia Appendix 1). They were then asked if they had been aware of doctor-rating websites before this survey. A total of 128 of 844 (15.2%) of respondents said that they had been aware of doctor-rating websites previously, and 716 (84.8%) said they had not.

The effect of 11 demographic variables (in Table 1) as well as 2 other relevant variables (1) being aware of the option to give feedback in general about GPs, and (2) having given feedback about GPs in the past were explored on the awareness of doctor-rating websites using chi-square tests and binomial logistic regression. The following 3 variables (Textbox 2) were found to be significant (see Table 3).

Qualifications and income were predictors for the awareness of the option to leave feedback using any method but were not found to be predictors for the awareness of doctor-rating websites.

Which, If Any, of the Following Doctor-Rating Websites Are You Aware Of?

From the 128 of 844 (15.2%) respondents who were aware of doctor-rating websites, 54/128 (42.2%) said they were not aware of a specific website. In total, 61/128 (47.7%) were aware of NHS Choices feedback site, 20/128 (15.6%) were aware of Patient Opinion, 5/128 (3.9%) were aware of PrivateHealth, 1/128 (0.8%) were aware of iwantgreatcare, and 2/128 (1.6%) mentioned “other.” This means that from all the respondents, only 61/844 (7.2%) were aware of the NHS Choices feedback site, and 20/844 (2.4%) were aware of Patient Opinion.

Results on Past Usage of Online Rating Websites

Past Experience of Giving Feedback About General Practitioners Using Any Method

There were 161 of 844 (19.1%) respondents that said they had formally given feedback about the care they had received from a GP in the past, and 683/844 (80.9%) said they had not. Of those who had given feedback formally in the past, 94/161 (58.4%) had given it directly to the GP, and 57/161 (35.4%) had given it to the GP practice. The remaining 10/161 (6.2%) had given it to an external organization.

The effect of 11 demographic variables (Table 1) on whether someone had given feedback in the past about their experience of receiving care from a GP was explored using chi-square tests and binomial logistic regression. There were 2 variables (Textbox 3) found to be significant (also see Table 4).

Past Usage of Doctor-Rating Websites for Any Purpose

Respondents who were aware of doctor-rating websites were asked if they had used a doctor-rating website before. Nineteen out of 128 (14.8%) had done so in the past, and the remaining 108/128 (84.4%) had not. This means that in total, from all the respondents, only 15/844 (1.8%) had used a doctor-rating website before. Given the amount the NHS and other external organizations have invested in establishing OPF websites, and the popularity of other rating websites like TripAdvisor, the very low level of usage at 15 is surprising.

The effect of 11 demographic variables (in Table 1) on the usage of doctor-rating websites was explored using Fisher’s exact test and binomial logistic regression. The variables ethnic origin (P=.043) and region (P=.041) as well as having searched the internet for health information previously (P=.007) were found using Fisher’s exact test to be significant on the usage of doctor-rating websites. The combined effect of all variables was investigated using binomial logistic regression; however, none of the variables were found to be significant (P>.05). Thus, it would seem that while having searched for health information in the past was found to be a predictor for the awareness of doctor-rating website and future consideration of using doctor-rating websites; it is not a predictor for usage.

Table 1. The 11 demographic characteristics of the respondents of the questionnaire (n=844).

Demographic characteristics	Respondents, n (%)		Difference between unweighted and weighted data, %
	Unweighted data	Weighted data	
Gender			
Male	433 (51.3)	413 (48.9)	–2.4
Female	411 (48.7)	431 (51.1)	+2.4
Age (years)			
15-24	150 (17.8)	132 (15.7)	–2.1
25-34	112 (13.3)	142 (16.8)	+3.5
35-44	116 (13.7)	134 (15.9)	+2.2
45-54	138 (16.4)	144 (17.1)	+0.7
55-59	58 (6.9)	51 (6.1)	–0.8
60-64	67 (7.9)	63 (7.4)	–0.5
65+	203 (24.1)	178 (21.0)	–3.0
Social grade[a]			
AB	191 (22.6)	231 (27.4)	+4.8
C1/C2	435 (51.5)	412 (48.8)	–2.7
D	124 (14.7)	129 (15.3)	+0.6
E	94 (11.1)	72 (8.6)	+2.5
Government office region			
East Midlands	56 (6.6)	73 (8.6)	+2.0
Eastern	71 (8.4)	94 (11.1)	+2.7
London	137 (16.2)	130 (15.5)	–0.7
North East	41 (4.9)	41 (4.9)	0.0
North West	126 (14.9)	111 (13.2)	–1.7
South East	111 (13.2)	137 (16.3)	+3.1
South West	100 (11.8)	86 (10.2)	–1.6
West Midlands	101 (12.0)	88 (10.4)	–1.6
Yorkshire and Humber	101 (12.0)	84 (9.9)	–2.1
Qualification			
GCSE/ O-LV/CSE/NVQ12[b]	215 (25.5)	212 (25.1)	–0.4
A-level or equivalent	168 (19.9)	160 (18.9)	–1.0
Bachelor/Master/PhD	234 (27.7)	264 (31.3)	+3.6
No formal qualification	168 (19.9)	150 (17.8)	–2.1
Other	59 (7.0)	59 (7.0)	0.0
Income (£)			
<11,499	102 (12.1)	88 (10.4)	–1.7
11,500-17,499	78 (9.2)	76 (9.0)	–0.2
17,500-24,999	47 (5.6)	45 (5.4)	–0.2
25,000-29,999	56 (6.6)	54 (6.4)	–0.2
30,000-39,999	63 (7.5)	68 (8.0)	+0.5
40,000-49,999	49 (5.8)	54 (6.4)	+0.6
50,000-74,999	66 (7.8)	86 (10.2)	+2.4

Demographic characteristics	Respondents, n (%)		Difference between unweighted and weighted data, %
	Unweighted data	Weighted data	
>75,000	35 (4.1)	44 (5.3)	+1.2
Don't know	158 (18.7)	153 (18.2)	–0.5
Refused	190 (22.5)	176 (20.8)	–1.7
Ethnicity			
White	710 (84.1)	723 (85.9)	+1.8
Non-white	134 (15.9)	118 (14.1)	–1.8
Internet access frequency			
Daily	657 (77.8)	679 (80.4)	+2.6
Weekly	67 (7.9)	62 (7.3)	–0.6
Monthly	14 (1.7)	12 (1.5)	–0.2
Never	106 (12.6)	91 (10.8)	–1.8
Have you ever used the internet to search for health information?			
Yes	434 (51.4)	458 (54.2)	+2.8
No	410 (48.6)	386 (45.8)	–2.8
Do you have a long-term health condition?			
Yes	241 (28.6)	222 (26.3)	–2.3
No	603 (71.4)	622 (73.7)	+2.3
Approximately how many General Practitioners are there in your current general practitioner surgery?			
1	31 (3.7)	29 (3.5)	–0.2
2-3	203 (24.1)	197 (23.3)	–0.8
4-5	265 (31.4)	268 (31.8)	+0.4
6-9	206 (24.4)	210 (24.9)	+0.5
>10	45 (5.3)	45 (5.3)	0.0
Don’t know	94 (11.1)	95 (11.2)	+0.1

[a]A: Higher managerial, administrative and professional; B: Intermediate managerial, administrative and professional; C1: Supervisory, clerical and junior managerial, administrative and professional; C2: Skilled manual workers; D: Semi-skilled and unskilled manual workers; E: State pensioners, casual and lowest grade workers, unemployed with state benefits only.

[b]GCSE: General Certificate of Secondary Education; O-LV: General Certificate of Education: Ordinary Level; CSE: Certificate of Secondary Education; NVQ: National Vocational Qualification.

Textbox 1. The 4 significant variables.

1. Income (£): This was found to be statistically significant (*P*=.003), and those with an income of £50,000-£74,999 had the highest odds and were 2.2 times more likely to be aware of the option to give feedback about their experience of care from a general practitioner (GP), in comparison to those whose income was below £11,499.
2. Qualification: This was found to be statistically significant (*P*=.002), and those with a graduate qualification had the highest odds and were also 2.2 times more likely to be aware than those with no formal qualifications.
3. The presence or absence of a long-term condition: This was found to be statistically significant (*P*=.004), and those who did have a long-term condition were 1.6 times more likely to be aware of the option to give feedback about a GP than those who did not have a long-term condition.
4. The number of GPs in the respondents’ surgery: This was also found to be statistically significant (*P*=.02), with those who were not aware of the number of GPs present in their surgery being the least likely (64.4%) to be aware of the option to give feedback about GPs, as compared with those who were aware that they had 1 GP in their surgery.

Table 2. Odds ratio adjusted for all the other variables for the effect of set demographic variables on the awareness of the option to give feedback about a general practitioner (n=844). The term "Ref" refers to the reference category (odds ratio of 1.000).

Variable	Odds ratio	95% CI
Income (£)[a]		
<11,499	Ref (1.000)	—
11,500-17,499	1.790	0.937-3.420
17,500-24,999	1.303	0.608-2.792
25,000-29,999	1.126	0.547-2.317
30,000-39,999	1.307	0.660-2.591
40,000-49,999	0.892	0.425-1.872
50,000-74,999[b]	2.211	1.131-4.320
>75,000	0.534	0.234-1.219
Don't know	0.789	0.436-1.429
Refused	0.826	0.472-1.445
Qualification[c]		
No formal qualification	Ref (1.000)	—
GCSE/O-level/CSE/NVQ[d]	1.020	0.628-1.659
A-level or equivalent	1.386	0.832-2.309
Degree/masters/PhD or equivalent[b]	2.197	1.350-3.575
Other	1.463	0.761-2.811
Long-term condition[e]		
No	Ref (1.000)	—
Yes[b]	1.631	1.166-2.283
Number of General Practitioners in the surgery[f]		
1	Ref (1.000)	—
2-3	0.902	0.389-2.090
4-5	0.899	0.392-2.065
6-9	0.867	0.372-2.018
>10	0.479	0.170-1.352
Don't know[b]	0.356	0.138-0.917

[a]*P*=.003

[b]*P*=.05

[c]*P*=.002

[d]GCSE: General Certificate of Secondary Education; O-LV: General Certificate of Education: Ordinary Level; CSE: Certificate of Secondary Education; NVQ: National Vocational Qualification.

[e]*P*=.004

[f]*P*=.019

Textbox 2. The 3 significant variables.

1. Age: This was found to be significant (*P*=.02), with those between the ages of 60-64 being 63% less likely to be aware of doctor-rating websites than those aged 35-44.
2. Those who had searched for health information on the internet in the past were 2.7 times more likely to be aware of doctor-rating websites than those who had not.
3. Also, those who were aware of the option to give feedback about a general practitioner using any method, were 5.6 times more likely to be aware of the existence of doctor-rating. websites than those who were not aware, suggesting that being aware of any method of giving feedback is a predictor for awareness of doctor-rating websites.

Table 3. Adjusted odds ratio for the effect of a set of demographics and 2 other variables on whether someone was aware of doctor-rating websites. Each odds ratio is adjusted for all the other variables in the table (n=844). The term "Ref" refers to reference category (odds ratio of 1.000).

Variable	Odds ratio	95% CI
Age (years)[a]		
15-24	0.425	0.181-1.000
25-34	1.442	0.753-2.762
35–44	Ref (1.000)	—
45-54	0.974	0.493-1.927
55–59	1.473	0.627-3.461
60–64[b]	0.366	0.127-1.057
>65	0.779	0.399-1.523
Past use of internet to search for health information[c]		
No	Ref (1.000)	—
Yes[b]	2.690	1.709-4.234
Awareness of the option to give feedback about general practitioners[c]		
No	Ref (1.000)	—
Yes[b]	5.632	3.631-8.737

[a]*P*=.02
[b]*P*=.05
[b]*P*<.001

Textbox 3. The 2 significant variables.

1. Gender was found to be statistically significant (*P*=.002), with female respondents almost twice as likely to have given feedback in the past than male respondents.
2. The presence or absence of a long-term health condition was found to be significant (*P*=.002), with those with a long-term health condition 1.8 times more likely to have given feedback about their experience of receiving care from a general practitioner in the past.

Table 4. Odds ratio adjusted for all the other variables in the table for the effect of a set of demographics on whether someone had given feedback about their experience of receiving care from a general practitioner in the past (n=844). The term "Ref" refers to reference category (odds ratio of 1.000).

Variable	Odds ratio	95% CI
Gender[a]		
Female	Ref (1.000)	—
Male	0.574	0.403-0.819
Long-term health condition		
No	Ref (1.000)	—
Yes	1.782	1.233-2.576

[a]*P*=.002

Textbox 4. The 7 significant variables.

1. Gender: This was found to be statistically significant (*P*=.01), with male respondents less likely to consider giving feedback in the future than females.
2. Age: This was also found to be statistically significant (*P*=.001), with those aged between 35-44, 55-59, and 60-64 around 2.5 times more likely to consider leaving feedback than those aged >65.
3. Long-term health condition: These were also twice more likely to consider leaving feedback than those did not have a health condition, as may be expected.
4. Who had used the internet in the past to search for health information: These were more than twice as likely to consider leaving feedback in the future than those who had not used the internet in the past to search for health information.
5. Number of general practitioners (GPs) in the respondents' surgery: This was also found to be significant with those who had 2-3 GPs in their surgery found to be 2.5 times more likely to consider leaving feedback than those who had just 1 GP in their surgery.
6. Qualification: This was also found to be statistically significant (*P*<.001), with those who had a graduate qualification being 4 times more likely to consider leaving feedback than those with no qualifications, and those with GCSEs or equivalent twice as likely to leave feedback than those with no qualifications.
7. Region: This was also found to be significant (*P*<.001), with those living in the North West, South East and Yorkshire and Humber, twice as likely to consider leaving feedback than those living in London, and those living in the North East 4.8 times more likely to consider leaving feedback than those living in London.

Past Usage of Doctor-Rating Websites for Giving Feedback About a General Practitioner

From the 19 respondents who had used a doctor-rating site before, 8/19 (42.1%) had used it to read a review for a doctor or hospital, 5/19 (26.3%) had used it to find a doctor or hospital, 4/19 (21.1%) had used it to review their experience of the NHS, and 3/19 (15.8%) had used it to give feedback about their experience of receiving care from a GP. Therefore, only 3/844 (0.4%) of the entire sample of respondents had used a doctor-rating website in the past to give feedback about their experience of receiving care from a GP.

From the 3 participants that left feedback on a doctor-rating website about a GP, 2 commented on a positive experience, and 1 commented on a negative experience. The reasons the 3 respondents gave for leaving feedback online was that they either wanted to let the GP know how much they appreciated the consultation or they believed sharing their experience would benefit the GP, or they wanted to comment on their treatment in general. No other reasons were cited.

Future Use of Online Rating Websites

Consideration of Giving Feedback in the Future Using Any Method

All respondents were asked whether they would consider giving feedback in the future about their experience of receiving care from a GP. A total of 638 of 844 (75.6%) respondents said they would consider giving feedback in the future, 214 (25.4%) said definitely, and 424 (50.0%) said possibly. A total of 199 (23.6%) said they would not consider giving feedback in the future, and 7 (0.8%) said they do not know.

Responses were first combined to form a bivariate variable of yes and no. The effect of 11 demographic variables (in Table 1) on consideration of giving feedback in the future was then explored using chi-square tests and binomial logistic regression. Seven variables were found to be statistically significant (Textbox 4 and Table 5).

Consideration of Future Use of Doctor-Rating Websites to Give Feedback About General Practitioners

A total of 18 of 844 (2.1%) respondents said they would consider using doctor-rating websites to give feedback about their experience of care from a GP (ie, a GP who is based in a surgery).

The effects of the 11 demographic variables (in Table 1) on the consideration of future use of doctor-rating websites was explored as well as the following additional variables: (1) awareness of doctor-rating websites, (2) past use of doctor-rating websites, (3) consideration of future use of doctor-rating websites for any purpose, and (4) consideration of giving feedback in the future about a GP. After using chi-square tests and binomial logistic regression, only past use of internet to search for health information remained significant (*P*=.007; please see Textbox 5 and Table 6).

Public Preference on Mode of Feedback

All respondents who said they would consider giving feedback in the future about a GP (n=776) were asked which mode they would most prefer using to give feedback about their experience with a GP, for both negative and positive feedback. They were provided with a list of 15 methods and were first asked to select the top 3 most preferred ways (or modes) to leave feedback and then their main preference. The complete sets of results are provided in Multimedia Appendix 2.

In summary (see Figure 1), the main preferences of respondents for giving feedback about their experience with a GP was (1) giving feedback directly to the GP where 397/776 (51.2%) selected this for positive feedback, and 348 (44.8%) for negative feedback, (2) giving feedback to the GP surgery manager where 84 (10.8%) for positive, and 123 (15.9%) for negative, (3) filling in a feedback form at the surgery or on the practice's website where 115 (14.8%) for positive, and 130 (16.8%) for negative, (4) posting feedback on a public website where 33 (4.3%) for positive, 36 (4.6%) for negative, and (5) giving feedback through an app where 29 (3.7%) for positive, and 33 (4.3%) for negative.

Table 5. Adjusted odds ratio for all the other variables in the table for the effect of a set of demographics on whether someone will consider giving feedback in the future about a general practitioner (n=844). The term "Ref" refers to reference category (odds ratio of 1.000).

Variable	Odds ratio	95% CI
Gender[a]		
Female	Ref (1.000)	—
Male[b]	0.630	0.438-0.906
Age (years)[c]		
15-24	0.866	0.457-1.638
25-34	1.607	0.833-3.102
35-44[b]	2.617	1.328-5.156
45-54	0.864	0.475-1.570
55-59[b]	2.555	0.992-6.578
60-64[b]	2.483	1.071-5.754
>65	Ref (1.000)	—
Region[c]		
London	Ref (1.000)	—
East Midlands	0.584	0.284-1.200
Eastern[b]	0.825	0.428-1.590
North East[b]	4.823	1.489-15.628
North West[b]	2.330	1.167-4.649
South East[b]	2.448	1.178-5.084
South West	2.298	1.055-5.003
West Midlands	0.979	0.512-1.870
Yorks and Humber[b]	2.357	1.093-5.082
Qualifications[c]		
No Formal Qualifications	Ref (1.000)	—
GCSE/O-Level/CSE/NVQ[d]	2.126	1.238-3.650
A-Level or Equivalent (=NVQ3)	1.714	0.952-3.084
Bachelors/Masters/PhD Or Equivalent[b]	4.086	2.287-7.298
Other[b]	2.649	1.166-6.019
Past use of internet to search for health information[c]		
No	Ref (1.000)	—
Yes[b]	2.392	1.624-3.524
Long-term health condition[e]		
No	Ref (1.000)	—
Yes[b]	2.078 (1.000)	1.257-3.433
No. of General Practitioners in surgery[f]		
1	Ref (1.000)	—
2-3[b]	2.511	1.034-6.097
4-5	2.010	0.823-4.911

Variable	Odds ratio	95% CI
6-9	2.275	0.894-5.794
>10	2.318	0.648-8.286
Don't know	0.759	0.292-1.975

[a]*P*=.013

[b]*P*=.05

[c]*P*<.001

[d]GCSE: General Certificate of Secondary Education; O-LV: General Certificate of Education: Ordinary Level; CSE: Certificate of Secondary Education; NVQ: National Vocational Qualification.

[e]*P*=.004

[f]*P*=.002

Textbox 5. The 1 significant variable after logistic regression.

Only past use of internet to search for health information remained significant (*P*=.007), with those who had used the internet to search for health information in the past being 1.6 times more likely to consider using doctor-rating websites to give feedback about a general practitioner (GP), than those who had not previously used the internet to search for health information. This suggests that existing engagement and interest in health, as well as being an indicator for patient awareness of doctor-rating websites (as mentioned earlier), is also an indicator for patient intention to use doctor-rating websites in the future to give feedback about GPs.

Table 6. Logistic regression (odds ratio) showing the effect of past use of the internet to search for health information on whether someone would consider using a doctor-rating website to give feedback about a GP (n=844). The term "Ref" refers to reference category (odds ratio of 1.000).

Variable	Odds ratio	95% CI
Internet to search for health information[a]		
No	Ref (1.000)	—
Yes[b]	1.649	1.144-2.376

[a]*P*=.007

[b]*P*=.05

Figure 1. Summary of respondents' (n=776) main preference for giving feedback about their experience of receiving care from a general practitioner (GP).

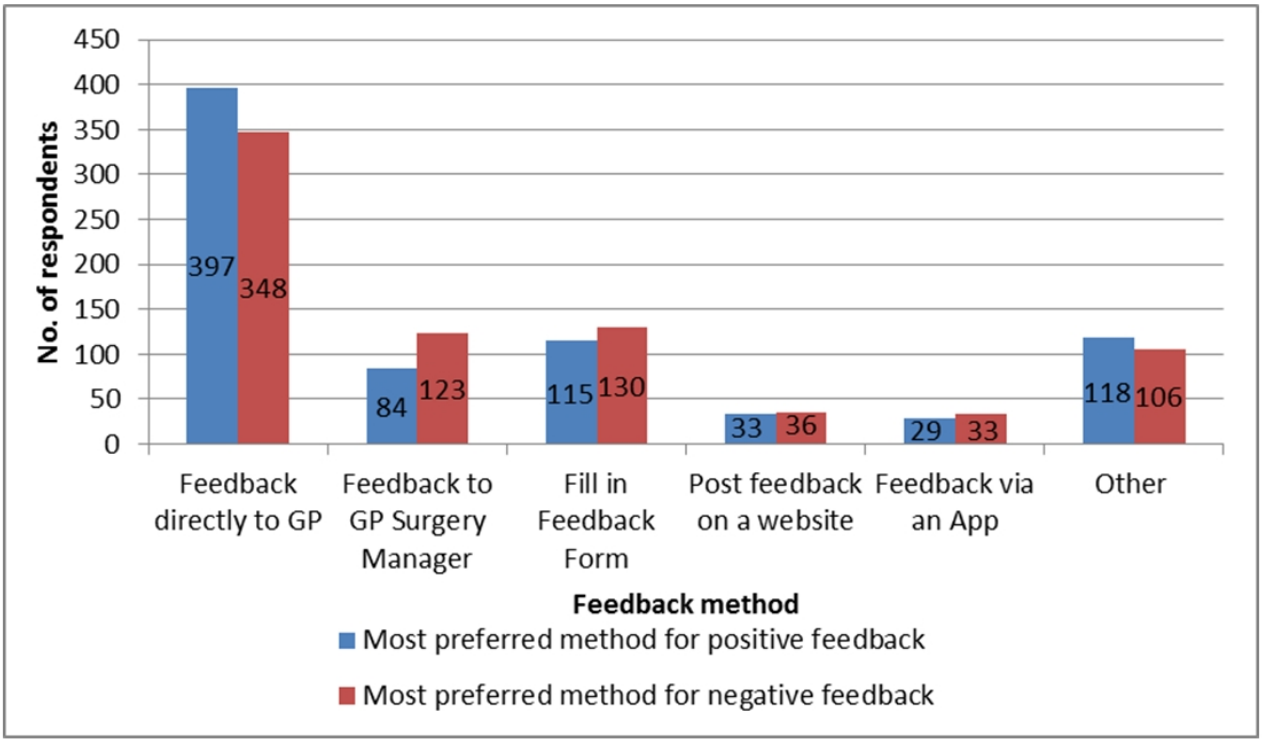


Table 7. Summary of key results from all respondents (n=844) relating to awareness, past usage and future consideration of giving feedback about experience with receiving care from a general practitioner (GP). Additional details are presented in the subsections.

Characteristic	Awareness		Past Use		Future Consideration	
	Any method	Doctor-rating websites	Any method	Doctor-rating websites	Any method	Doctor-rating websites
Positive (yes), n (%)	329 (39.0)	126 (15.0)	160 (19.0)	3 (0.4)	641 (75.9)	151 (17.9)
Gender	—	—	Y[a]	—	Y	—
Age	—	Y	—	—	Y	—
Social grade	—	—	—	—	—	—
Region	—	—	—	—	Y	—
Qualification	Y	—	—	—	Y	—
Income	Y	—	—	—	—	—
Ethnicity	—	—	—	—	—	—
Internet access frequency	—	—	—	—	—	—
Past use of internet to search for health information	—	Y	—	—	Y	Y
Presence or absence of a long-term health condition	Y	—	Y	—	Y	—
Number of GPs in surgery per local health center	Y	—	—	—	Y	—

[a]Y: significant using binomial logistic regression.

It is interesting to note that although results in the previous sub-section indicate that 150/844 (17.0%) of all respondents would consider using doctor-rating websites (both NHS and independent websites) in the future to give feedback about GPs. Only 36 respondents selected a doctor-rating website as their most preferred method to leave negative feedback about GPs. This corresponds to 36/776 (4.6%) of all those who would consider giving feedback, and 36/844 (4.3%) of all respondents. The overwhelming preference for leaving feedback with GPs or the GP surgery correlates with earlier results that indicated that 151/161 (93.8%) of those who had left feedback for or about a GP in the past, had left it with the GP or GP practice.

Figure 1 also demonstrates that patients' most preferred method to give feedback varies depending on whether their feedback is about a negative or a positive experience. This suggests "patient feedback mode" is partially dependent upon the nature of the experience.

Summary of the Results

Table 7 summarizes the key results found in this study, as well as the demographic factors that were found to be significant on each of the key dependent variables. The table demonstrates that 128 of 844 (15.2%) respondents were aware of doctor-rating websites for giving feedback about GPs, in comparison to 326 (38.6%) of respondents who were aware of giving feedback using any of the methods. Similarly, 161 (19.1%) of respondents had given feedback about a GP in the past using any method, whereas only 3 (0.4%) had given feedback about a GP using a doctor-rating website. A total of 638 (75.6%) of respondents said they would consider giving feedback about a GP in the future (using any method); whereas only 150 (17.8%) of respondents said they would consider giving feedback in the future using doctor-rating websites.

Discussion

Public Awareness of Doctor-Rating Websites

The results suggest that based on a representative sample of 844 respondents, 15.2% (128/844) of the population in England is aware of the existence of doctor-rating websites to give feedback to a GP, whereas 38.6% (326/844) is aware that they can give feedback using any method. The level of awareness found in this study is in line with findings from a previous study by Galizzi et al [35] who found that 15% of their London-based respondents was aware of the existence of doctor-rating websites, although it was not clear for which purpose they were aware of such websites, and which specific websites they were aware of. However, they suggested that this indicated low awareness amongst the population in England.

The findings from the present study suggest that awareness of doctor-rating websites to give feedback about a GP, compared with awareness of the option to give feedback about a GP using any method, is not low. This is because almost half of those who are aware of the option to give feedback about a GP are aware of the existence of doctor-rating websites (for feedback on GPs). Despite this, 54/128 (42.2%) of those that were aware of doctor-rating websites were not aware of a specific website, only 61/844 (7.2%) respondents were aware of the NHS Choices feedback website, and 20/844 (2.4%) of Patient Opinion. This indicates that awareness of specific doctor-rating websites is low, which is surprising given that the NHS Choices feedback website is an official channel for patients in England to leave feedback about healthcare services (although it is unknown how well if at all, it is promoted to patients and the public).

Higher levels of awareness of doctor-rating websites were found outside of the UK, with the highest found in the USA at 65% by Hanauer et al [31], and in Germany, at 29.3% in 2012 [32]

and 32% in 2013 [18]. The higher levels of awareness in comparison to what was found in this study may be partially explained by the higher usage and popularity of private healthcare, the competitive nature of healthcare in both countries, and also what may appear to be a higher usage of internet for health seeking information (reported in one study in Germany at 68% [43] when compared to 54% found in this study). But there could also be a sampling effect, as the aforementioned studies were all conducted using online panel sampling. However, a recent study in Germany that used a cross-sectional random sample survey found awareness at 72.5% [21]

The results from the present study indicate that awareness of doctor-rating websites (unlike awareness of giving feedback to a GP in general), is not dependent on being wealthier, having better qualifications, having a long-term condition (and possibly using GP services more frequently) and knowing how many GPs practice in your surgery. Instead, age and having searched for health information in the past were found to be the only predictors for awareness of doctor-rating websites. Age was also found to be significant by Galizzi et al [35], and this they suggest is not surprising because elderly people use the internet less frequently. If a person has searched for health information in the past, this may suggest that: (1) they know how to use the internet (and may have access to it too), and (2) they are actively interested in their health. It is therefore not surprising that they are more likely to be aware of doctor-rating websites.

In London, Galizzi et al [35] found that as well as age, ethnicity was significant for awareness, with white respondents less likely to be aware of these websites; however, this was not found in this study. In Germany, Emmert et al [18] found that differences in age group were not statistically significant, and neither was education, employment, internet use, and health status. However, unlike this study, they found that female respondents were more likely to be aware of doctor-rating websites, as well as those widowed, and those with higher health care utilization. In this study, female respondents were found to be more likely than male respondents to have given feedback in the past using any method.

Public Usage of Doctor-Rating Websites

Based on the present sample, 19.1% (161/844) of the population in England has given feedback in the past using any method, whereas only 0.4% (3/844)of the population have given feedback using doctor-rating websites, which is significantly lower. The level of use of doctor-rating websites to specifically give feedback or review GPs in England had not been explored in previous studies; however, Galizzi et al [35] did explore usage of doctor-rating websites and found that 3% of their Londoners' sample (n=200) had used doctor-rating websites, although again it was not evident for which purpose. This is similar to the finding of this study that 1.8% (15/844) of the population had used a doctor-rating website before for any of the purposes. The low level of usage indicates that patients are not using doctor-rating websites, especially not to give feedback about GPs. This is surprising given that the NHS recently spent £1.25M piloting a new doctor-rating website called CareConnect [44].

Outside of the UK, usage of doctor-rating websites was found to be much higher. A recent study in the USA (in Rochester Minnesota) reported that 16% of the surveyed sample had used a doctor-rating website, and 3% had used it to give feedback [37]. This was different to another study in the USA, conducted in 2012, which reported general usage at 23%, and found that 6% of parents had left ratings for doctors online [45]. The difference may be partly because the latter study used an online panel as the survey mode, and the former study used a written questionnaire, or that the former was conducted in a single healthcare setting.

Usage of doctor-rating websites was also high in Germany. In 2013, 25% of the population had used a doctor-rating website to search for a doctor, and 11% to leave feedback or ratings [18]. Similarly, Terlutter et al [32] discovered in Germany (in 2012) that 26% of the population had used a doctor-rating website before, although it was not clear for which purpose. More recently, in 2017, researchers found usage of doctor-rating websites in Germany at 43.6% [21] In Austria, researchers conducted an experimental study based on a convenient sample and found that 47% of respondents had used a doctor-rating website, and 6% had used it to leave feedback [46]. The difference in results may be due to regional differences in the diffusion of doctor-rating websites and the adoption of the internet for seeking health information. However, there may also be a sampling effect, because many of the studies outside of the UK (with the exception of the one conducted by Burkle and Keegan [37] and McLennan et al [21]) used an online panel as their sample population. The use of online sampling may have affected results, because those who are online, and had used the internet to search for health information, may be more likely to be aware of and use doctor-rating websites than those that had not, as results from this present study suggest.

In the USA and Germany, academics found various predictors for usage of doctor-rating websites, such as the presence of a long-term condition, advanced education, age, and gender [18,32,37,45]. Predictors for the usage of doctor-rating websites for feedback about GPs could not be computed in this study because only 0.4% (3/844) of respondents had used a doctor-rating website for that purpose. However, the results do indicate however that female respondents and those with long term health conditions are significantly more likely to have given feedback in the past to a GP (using any method). Those with long-term health conditions tend to use GP services more than those who do not have a long-term health condition, and so it is not surprising that they are more likely to leave feedback.

Future Use of Doctor-Rating Websites

Although the present study suggests that 75.6% (638/844) of the population in England would consider giving feedback in the future to a GP using any method, only 17.8% (150/844) would consider giving feedback in the future to a GP on a doctor-rating website. This suggests that more than half of respondents would consider giving feedback to a GP but not on a doctor-rating website. Similarly, 33.1% (279/844) of the population would consider using doctor-rating websites but not to leave feedback for a GP. This, as well as the 0.4% (3/844) past usage of doctor-rating websites, and only 4.3% (33/776)

to 4.7% (36/776) selecting doctor-rating websites as their most preferred feedback method, questions whether doctor-rating websites are wanted or needed by the public for leaving feedback about GPs.

The only significant predictor for the future use of doctor-rating websites for giving feedback about GPs was the past use of the internet to search for health information, with those that had were found to be 1.6 times more likely to consider using doctor-rating websites to give feedback about a GP than those that had not. This predictor is not surprising given it indicates an active interest in one's health as well as familiarity with the internet. What was surprising was the absence of 6 other predictors which were found to be significant for the future consideration of using any method to leave feedback about GPs. These predictors indicated that those that are either female, younger in age, have a long-term health condition, have higher qualifications, have more GPs in their surgery, or live outside of London are much more likely to consider leaving feedback about a GP using any method. This could be seen as a positive suggestion that doctor-rating websites, unlike other feedback methods, may span across the age, social and regional divide, and appeal to everyone who takes an active interest in their health and is familiar with the internet to pursue that interest. This appears to support Bardach et al's [26] argument that OPF websites would collect feedback from those patients who would not normally give feedback. For consideration of using doctor-rating websites in the future for any purpose (and not just giving feedback about a GP), in addition to past use of the internet to search for health information, the respondent's age and internet consumption were also found to be significant predictors. This is in contrast to Galizzi et al's [35] findings with Londoners in which income, ethnicity, and the doctor-patient relationship were the significant predictors for future intention to use doctor-rating websites for any purpose.

Public Preference on Mode of Feedback

The results suggest that there is no one most preferred way for patients to leave feedback about a GP, and this was also found by Patel et al However, like the results of Patel et al [47], the present study also found rather surprisingly that almost half of those who would consider leaving feedback for a GP would prefer to give feedback directly to the GP, even when it is negative feedback. Furthermore, the 2 major reasons for choosing 1 mode of feedback over another were ease and convenience, followed by the method being a direct way of giving feedback (and Patel et al [47] found that the latter was so that patient feedback reaches the GP and is used by the GP for improvement purposes). These are interesting findings because currently there is little formal provision in general practice in England to give feedback directly to the GP.

Current formal provisions for leaving feedback about GPs in the NHS also include the NHS Friends and Family Test card, which is a paper-based feedback form that is used in most GP practices in England [48], and the GP Patient Survey [48]. The paper-based feedback form was only selected by 10% of respondents (who would consider leaving feedback in the future) as their most preferred method for leaving negative feedback for a GP. Similarly, use of OPF websites to report a negative experience was selected as the main preference by only 5% of respondents. In contrast, 45% of patients' most preferred method to leave negative feedback was directly with the practice and 16% directly with the practice manager. The vast majority of patients (94%) who had given feedback in the past had given it directly to the GP or practice.

These results as well as others suggest that current methods available in general practice to leave feedback are on the whole not the most preferred methods for patients. Therefore, GP practices and the NHS need to consider alternative ways and methods to collect feedback. For example, giving patients the option to send feedback through email, which was selected by 12% of respondents (who would consider leaving feedback in the future) as their most preferred method. This also questions the value of OPF websites and questions whether patients in England want or need these types of websites to leave feedback about GPs.

Although preference for leaving feedback online was minimal, one of the interesting findings from the results was that more people prefer to leave feedback online on a private feedback form on the GP surgery website, rather than leaving it on a (public) doctor-rating website. Similarly, although more people preferred to give feedback directly to the GP in person or telephone, in comparison to writing a letter, more respondents preferred to use email to send feedback. Furthermore, an app was found to be almost the same in popularity as leaving feedback on a doctor-rating website, although again the main preference was to use an app that would give the feedback to the GP surgery directly rather than an app that would publish the feedback online. These findings support the notion that many patients prefer to give feedback directly to the GP and practice rather than leaving feedback in the public domain, and these alternative modes of leaving feedback need to be taken into consideration by GPs and GP practices in England, if they want to engage and increase the volume of patient experience feedback.

Strengths and Limitations

The strength of this study lay in its use of a well-validated mixed methods population questionnaire whose aim was to measure representative views of the public on giving feedback about GPs on OPF websites, within the context of other feedback mechanisms. Nevertheless, this study did have several limitations.

Firstly, the sampling method used—a random location quota sampling—was not a random sample, and although the data was weighted so that it would be a representative sample of the population in England, the sample may still contain biases, and claiming generalizability (external validity) across the whole population in England could be questioned. However, given that it was not feasible to get a random sample of the population in England, this was as close as possible to a true representative sample and very little correction of the results was needed to make them in line with The National Readership Survey (NRS), which uses random probability sampling. The interviews were also conducted face-to-face, which meant that there was very little risk of respondents misunderstanding the questions, and

there was a lower risk of premature termination, as interviewers could keep respondents motivated.

Secondly, although the questionnaire had strong internal validity, the fieldwork was conducted by 155 interviewers from Ipsos MORI, and not the authors, and this could be a potential weakness. Nevertheless, the interviewers were all experienced professional interviewers who were trained by Ipsos MORI and given the same very specific instructions. A validation procedure on the fieldwork was also conducted to ensure that interviewers had interviewed respondents as expected.

Thirdly, the results of this study question the value of providing OPF websites in England to give feedback about GPs; however, this study did not explore patients' views on OPF websites for choice, an issue that was outside the scope of this study. Although both giving feedback and patient choice are highly connected (because if patients do not give feedback or reviews online, other patients will not have these patient reviews to choose from), they are distinctly different as actions. The results when reported in this study make clear that they are specifically about giving feedback on GPs only.

Fourthly, this study focused on primary care and GPs only in England. The results may have been different if the study focused on other healthcare professionals such as surgeons, or secondary care.

Implications for Practice

The results from this study strongly suggest that GPs, GP practices, the NHS, and feedback website providers should consider alternative mechanisms to collect patient feedback in general practice, instead of relying primarily on the NHS Family and Friends Test card and online patient feedback websites as a day-to-day feedback method. In particular, direct methods to give feedback to the GP or the GP practice (digital or non-digital) are most used and preferred by patients, such as face-to-face feedback, email, telephone, and private feedback forms on the GP practice website. Therefore, we recommend the NHS to channel its investment and resources towards providing more direct and private feedback methods in general practice (such as opportunities for face-to-face feedback, email-based feedback, and Web-based private feedback forms), as these are much more likely to be used currently by the majority of patients in England. We also recommend that when online feedback is presented to other patients for choice", the feedback must be part of a collection of measures including patient feedback collected using other methods, and other measures such as the clinical competency of the GP, findings from the Care Quality Commission report, and safety results. Other recommendations for OPF providers and GPs and GP practices can be found in Multimedia Appendix 3 and Multimedia Appendix 4 respectively, where we also highlight what the NHS and other OPF providers can do to increase patient use of OPF websites.

Conclusions

This is the first piece of nationally representative research that has explored patients' awareness and usage of OPF websites within the context of other feedback mechanisms available in general practice in England, and to date, in our knowledge, the largest and most robust study conducted with patients about doctor-rating websites.

Given the popularity, acceptance, and usage of consumer rating websites such as Trip Advisor, coupled with the increasing emphasis on PPI and patient experience in the NHS, and the millions of pounds investment into OPF websites by the NHS [44], it is surprising that this study (alongside Patel et al [36]), unlike previous academic work on online rating websites, questions whether patients and carers really want or need OPF to give feedback about GPs in England.

This is because the findings indicate that although awareness is not so poor of doctor-rating websites when compared to awareness of giving feedback in general, past usage is extremely uncommon at 0.4% for feedback about GPs, and so is future consideration to use doctor-rating websites for giving feedback about GPs 82.0% of the public indicated that they will not consider using doctor-rating websites to give feedback in the future; although a further 32.9% of the population would consider using doctor-rating websites but not to leave feedback for a GP. Furthermore, only 4.0%-5.0% of those who would consider leaving feedback in the future selected doctor-rating websites as their most preferred method to leave feedback about a GP.

This, as well as the different predictors found for awareness, usage, and future consideration to use OPF websites, all appear to suggest that (1) OPF websites may not be an effective channel for collecting feedback on patient experience in general practice (and hence the NHS should provide alternative methods of collecting feedback), and (2) feedback on OPF websites is not likely to be representative of the patient experience in the near future. Although this may not be a pertinent problem for GPs and GP providers using the patient experience data for improvement (because improvement even based on 1 piece of patient feedback could potentially be useful), fundamentally it is a huge problem for the use of OPF for selection (ie, patient choice), and for monitoring. This is because the results suggest that OPF is biased because it is not representative of patient experience, and therefore patients using OPF for choice of healthcare provider are basing their choice on biased and unrepresentative data, challenging strongly the popular notion that OPF is useful for patient choice, as advocated both by academics [9,11,14] and the NHS [49,50]. Furthermore, the findings appear to contradict Greaves et al's [14] observation of associations between NHS Choices general practice ratings and patient experience measures, thus strongly questioning the usefulness of OPF as a measure of quality in health care.

Nevertheless, the findings do suggest that OPF websites fulfill a feedback gap" for a very small number of patients, and appear to support the argument that some patients, who would not normally give feedback using other methods, would leave feedback on OPF websites. Therefore, this may suggest that OPF websites could be used to improve patient experience, as feedback can be collected from those patients who may not give feedback using other channels, as long as GP Practices were willing to use OPF for improvement purposes.

Acknowledgments

We would like to thank all the participants who took part in this study, and IPSOS Mori for their assistance with the fieldwork, in particular, Sarah Shephard, Penny Bowden, and Ross Connell. The research was funded by the Engineering and Physical Sciences Research Council (EPSRC) under the Participation in Healthcare Environment Engineering Programme (#EP/H022031/1). The underlying research data from this study cannot be made public as participants' consent was not taken to place the raw data in the public domain.

Authors' Contributions

This research was designed, conducted, and written by the first author (SP) as part of her doctoral research at the University of Warwick. The remaining authors reviewed the study design, questionnaire, and manuscript. The second author (RC) was employed at the time by the University of Warwick.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire (final validated version) used in this study.

[PDF File (Adobe PDF File), 223KB - jmir_v20i7e243_app1.pdf]

Multimedia Appendix 2

Complete set of results for the question: In which, of the following ways, if any, would you prefer to give feedback about a GP?

[PDF File (Adobe PDF File), 156KB - jmir_v20i7e243_app2.pdf]

Multimedia Appendix 3

List of recommendations for online patient feedback website providers and owners.

[PDF File (Adobe PDF File), 25KB - jmir_v20i7e243_app3.pdf]

Multimedia Appendix 4

List of recommendations for GPs and GP Practices.

[PDF File (Adobe PDF File), 25KB - jmir_v20i7e243_app4.pdf]

References

1. Office for National Statistics. 2015. Internet Access - Households and Individuals URL: http://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2015-08-06 [WebCite Cache ID 6zqXeNqxW]
2. Schuckert M, Liu X, Law R. Insights into Suspicious Online Ratings: Direct Evidence from TripAdvisor. Asia Pacific J Tour Res [Internet] 2016;21(3):259-272. [doi: 10.1080/10941665.2015.1029954]
3. Li H, Ye Q, Law R. Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis. Asia Pacific J Tour Res 2012;(April 2015)? 2013 Jul 01:19.
4. Filieri R. What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. Journal of Business Research 2015 Jun;68(6):1261-1270. [doi: 10.1016/j.jbusres.2014.11.006]
5. Forster R, Gabe J. Voice or Choice? Patient and Public Involvement in the National Health Service in England under New Labour. Int J Heal Serv Internet;38(2)?356 2008 Jan:333 [FREE Full text]
6. Baggott R. A Funny Thing Happened on the Way to the Forum? Reforming Patient and Public Involvement in the NHS in England. Public Administration 2005 Aug;83(3):533-551. [doi: 10.1111/j.0033-3298.2005.00461.x]
7. Coulter A, Angela. Engaging Patients in Healthcare. UK: McGraw-Hill; 2018.
8. Department of Health. London: Department of Health; 2015. The Handbook to the NHS Constitution [Internet] URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/474450/NHS_Constitution_Handbook_v2.pdf [WebCite Cache ID 6zqXZyHl0]

9. Lagu T, Goff SL, Hannon NS, Shatz A, Lindenauer PK. A mixed-methods analysis of patient reviews of hospital care in England: implications for public reporting of health care quality data in the United States. Jt Comm J Qual Patient Saf 2013 Jan;39(1):7-15. [Medline: 23367647]
10. Greaves F, Millett C, Nuki P. England's Experience incorporating “anecdotal” reports from consumers into their national reporting system: lessons for the United States of what to do or not to do? Med Care Res Rev 2014 Oct;71(5 Suppl):65S-80S. [doi: 10.1177/1077558714535470] [Medline: 24836765]
11. Trigg L. Patients' opinions of health care providers for supporting choice and quality improvement. J Health Serv Res Policy 2011 Apr;16(2):102-107. [doi: 10.1258/jhsrp.2010.010010] [Medline: 20870690]
12. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
13. Greaves F, Pape U, King D, Darzi A, Majeed A, Wachter R. Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf Internet;(April) 2012 Apr 20:a [FREE Full text]
14. Greaves F, Pape UJ, Lee H, Smith DM, 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;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
15. Hughes S, Cohen D. Can online consumers contribute to drug knowledge? A mixed-methods comparison of consumer-generated and professionally controlled psychotropic medication information on the internet. J Med Internet Res 2011;13(3):e53 [FREE Full text] [doi: 10.2196/jmir.1716] [Medline: 21807607]
16. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
17. 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]
18. 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;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
19. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating Website: an analysis of the ratings. J Med Internet Res 2015 Jan 12;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
20. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
21. 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 Nov 09;19(11):e387 [FREE Full text] [doi: 10.2196/jmir.7581] [Medline: 29122739]
22. Adams S. Post-Panoptic Surveillance through Healthcare Rating Sites: Who's watching whom? Information, Communication & Society 2013 Mar;16(2):215-235 [FREE Full text] [doi: 10.1080/1369118X.2012.701657]
23. 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;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
24. Rothenfluh F, Schulz PJ. 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 May 01;19(5):e127 [FREE Full text] [doi: 10.2196/jmir.6875] [Medline: 28461285]
25. 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]
26. Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
27. 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]
28. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012 Apr 10;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
29. Verhoef LM, Van DBTH, Engelen LJLPG, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
30. Bidmon S, Terlutter R, Röttl J. What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. J Med Internet Res 2014 Jun 11;16(6):e148 [FREE Full text] [doi: 10.2196/jmir.3122] [Medline: 24918859]

31. 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]
32. 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;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
33. 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;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
34. The Nuffield Trust. 2013. Rating providers for quality: a policy worth pursuing? URL: https://www.nuffieldtrust.org.uk/research/rating-providers-for-quality-a-policy-worth-pursuing [WebCite Cache ID 6aPkLR7vx]
35. Galizzi MM, Miraldo M, Stavropoulou C, Desai M, Jayatunga W, Joshi M, et al. Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. BMJ Open 2012;2(6):1-10 [FREE Full text] [doi: 10.1136/bmjopen-2012-001493] [Medline: 23148340]
36. Patel S, Cain R, Neailey K, Hooberman L. Exploring Patients' Views Toward Giving Web-Based Feedback and Ratings to General Practitioners in England: A Qualitative Descriptive Study. J Med Internet Res 2016 Aug 05;18(8):e217 [FREE Full text] [doi: 10.2196/jmir.5865] [Medline: 27496366]
37. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. BMC Health Serv Res 2015;15:416 [FREE Full text] [doi: 10.1186/s12913-015-1099-2] [Medline: 26410383]
38. Pennay D. Introducing the Total Survey Error (TSE) framework. 2014 Presented at: ACSPRI Soc Sci Methodol Conference; 2014; North America p. 1 URL: https://conference.acspri.org.au/index.php/conf/conference2014/paper/view/768/58
39. Gideon L. Handbook of Survey Methodology for the Social Sciences. New York: Springer Science & Business Media; 2012.
40. Field A, Andy. Discovering Statistics Using IBM SPSS Statistics. London: SAGE PUBN; 2018.
41. Marston L. Introductory Statistics for Health and Nursing Using SPSS. United Kingdom: Sage Publications Ltd; 2018.
42. Peacock J, Kerry S. Presenting Medical Statistics from Proposal to Publication. USA: Oxford University Press; 2006.
43. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015 Jan 29;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
44. Shah S. 25m on now-abandoned Care Connect service. NHS spent £1 URL: http://www.computing.co.uk/ctg/news/2434821/nhs-spent-gbp125m-on-now-abandoned-care-connect-service# [accessed 2016-05-12] [WebCite Cache ID 70pWrRGVU]
45. Hanauer D, Zheng K, Singer D, Gebremariam A, Davis M. Parental awareness and use of online physician rating sites. Pediatrics 2014 Oct;134(4):e966-e975 [FREE Full text] [doi: 10.1542/peds.2014-0681] [Medline: 25246629]
46. Grabner-Kräuter S, Waiguny M. Insights into the impact of online physician reviews on patients' decision making: randomized experiment. J Med Internet Res 2015 Apr 09;17(4):e93 [FREE Full text] [doi: 10.2196/jmir.3991] [Medline: 25862516]
47. Patel S, Cain R, Neailey K, Hooberman L. Exploring Patients' Views Toward Giving Web-Based Feedback and Ratings to General Practitioners in England: A Qualitative Descriptive Study. J Med Internet Res 2016 Dec 05;18(8):e217 [FREE Full text] [doi: 10.2196/jmir.5865] [Medline: 27496366]
48. Gillam S, Newbould J. Patient participation groups in general practice: what are they for, where are they going? BMJ 2016 Feb 18:i673 [FREE Full text] [doi: 10.1136/bmj.i673]
49. Lansley A. Department of Health. 2012. Response to NHS Future Forum's second report Internet URL: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_132088.pdf [WebCite Cache ID 6zqXncV2I]
50. NHS Choices website. Patient choice of GP practices - The NHS in England - NHS Choices Internet URL: http://www.nhs.uk/NHSEngland/AboutNHSservices/doctors/Pages/patient-choice-GP-practices.aspx [accessed 2016-05-11] [WebCite Cache ID 6zqXwL2gt]

Abbreviations

GP: general practitioner
NHS: National Health Service
NRS: National Readership Survey
OPF: Online patient feedback
PPI: patient and public involvement

Edited by T Kool; submitted 20.01.18; peer-reviewed by A Davey, R Terlutter, F Rothenfluh, S McLennan; comments to author 10.03.18; revised version received 24.04.18; accepted 05.06.18; published 25.07.18.

Please cite as:
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;20(7):e243
URL: http://www.jmir.org/2018/7/e243/
doi:10.2196/jmir.9523
PMID:30045831

Original Paper

Accuracy of Internet-Based Patient Self-Report of Postdischarge Health Care Utilization and Complications Following Orthopedic Procedures: Observational Cohort Study

Benjamin I Rosner[1,2], MD, PhD; Marc Gottlieb[3], MPA; William N Anderson, PhD

[1]HealthLoop Inc, Mountain View, CA, United States

[2]Department of Hospital Medicine, Kaiser Permanente, Santa Clara, CA, United States

[3]Anthem Inc, Atlanta, GA, United States

Corresponding Author:
Benjamin I Rosner, MD, PhD
HealthLoop Inc
605 Ellis St #100
Mountain View, CA,
United States
Phone: 1 408 418 0998
Fax: 1 888 972 8929
Email: ben@healthloop.com

Abstract

Background: The accuracy of patient self-report of health care utilization and complications has yet to be determined. If patients are accurate and engaged self-reporters, collecting this information in a manner that is temporally proximate to the health care utilization events themselves may prove valuable to health care organizations undertaking quality improvement initiatives for which such data are often unavailable.

Objective: The objective of this study was to measure the accuracy of patient self-report of health care utilization and complications in the 90 days following orthopedic procedures using an automated digital patient engagement platform.

Methods: We conducted a multicenter real-world observational cohort study across 10 orthopedic practices in California and Nevada. A total of 371 Anthem members with claims data meeting inclusion criteria who had undergone orthopedic procedures between March 1, 2015, and July 1, 2016, at participating practices already routinely using an automated digital patient engagement platform for asynchronous remote guidance and telemonitoring were sent surveys through the platform (in addition to the other materials being provided to them through the platform) regarding 90-day postencounter health care utilization and complications. Their self-reports to structured survey questions of health care utilization and complications were compared to claims data as a reference.

Results: The mean age of the 371 survey recipients was 56.5 (SD 15.7) years, 48.8% (181/371) of whom were female; 285 individuals who responded to 1 or more survey questions had a mean age of 56.9 (SD 15.4) years and a 49.5% (141/285) female distribution. There were no significant differences in demographics or event prevalence rates between responders and nonresponders. With an overall survey completion rate of 76.8% (285/371), patients were found to have accuracy of self-report characterized by a kappa of 0.80 and agreement of 0.99 and a kappa of 1.00 and agreement of 1.00 for 90-day hospital admissions and pulmonary embolism, respectively. Accuracy of self-report of 90-day emergency room/urgent care visits and of surgical site infection were characterized by a kappa of 0.45 and agreement of 0.96 and a kappa of 0.53 and agreement of 0.97, respectively. Accuracy for other complications such as deep vein thrombosis, hemorrhage, severe constipation, and fracture/dislocation was lower, influenced by low event prevalence rates within our sample.

Conclusions: In this multicenter observational cohort study using an automated internet-based digital patient engagement platform, we found that patients were most accurate self-reporters of 90-day hospital admissions and pulmonary embolism, followed by 90-day surgical site infection and emergency room/urgent care visits. They were less accurate for deep vein thrombosis and least accurate for hemorrhage, severe constipation, and fracture/dislocation. A total of 76.8% (285/371) of patients completed surveys without the need for clinical staff to collect responses, suggesting the acceptability to patients of internet-based survey dissemination from and collection by clinical teams. While our methods enabled detection of events outside of index institutions, assessment of accuracy of self-report for presence and absence of events and nonresponse bias analysis, low event prevalence

rates, particularly for several of the complications, limit the conclusions that may be drawn for some of the findings. Nevertheless, this investigation suggests the potential that engaging patients in self-report through such survey modalities may offer for the timely and accurate measurement of matters germane to health care organizations engaged in quality improvement efforts post discharge.

(J Med Internet Res 2018;20(7):e10405) doi:10.2196/10405

KEYWORDS

patient-generated health data; patient reported outcome measures; patient self-report; complications; utilization; patient readmission; emergency room; hospital economics

Introduction

Rates of health care utilization and complications post discharge are topics of increasing interest and value under both federal [1-3] and commercial [4,5] bundled payment and hospital readmission reduction initiatives [6-8]. However, accurate and timely measurement and reporting of these outcomes vary, in part due to limitations of the sources from which such data are derived, large variations in the ways in which they are measured [9], and the lag time between the capture of these events in reporting systems and their dissemination back to the very health care organizations at which the index encounters occurred [10]. In the rise of the era of the patient-as-partner-in-care, health care organizations engaged in quality improvement or those seeking to enhance performance in value-based reimbursement models may find patients to be uniquely valuable sources of information regarding rates of health care utilization and complications post discharge. However, for a feedback cycle between patients and health care organizations to be meaningful and useful in quality improvement, health care organizations need to (1) enable and engender patient participation in this cycle, (2) scale the low-burden dissemination of such surveys to patients, (3) attain high survey completion rates, and (4) feel confident that patients can be timely and accurate self-reporters.

This study aimed to understand whether patients in real-world clinical practice settings, surveyed at 90 days post encounter through an automated digital patient engagement (DPE) platform, are accurate self-reporters of real-time or near real-time readmissions, emergency room/urgent care (ERUC) usage, and postencounter complications.

Studies have been conducted investigating the accuracy of patient self-report on topics ranging from past medical history [11,12], surgical history [13], and diagnosis underlying the need for a given intervention [14]. Investigations have also described the accuracy of self-reported complications using either general practitioner surveys [15] or independent surgeon review of confirmatory studies [16] or medical records [17] as references. Some of these studies have been confounded by the methodological tautology of relying on patients to confirm their own self-reports [16,17], and others have been limited in their completeness by assessing accuracy only among patients reporting the presence of events without also assessing accuracy of those reporting the absence of events [15-17]—the latter being a cohort that is much larger when examining low prevalence events such as readmissions and complications, and arguably just as important from a quality improvement perspective.

Other studies have been prone to the potential for recall error, sometimes referred to as memory decay, caused by the lag time between when the event occurred and the relatively distant time at which the patient was later surveyed for such events [17,18]. Stability of accurate patient recall for such events appears to remain over at least 2 to 3 months [19,20] but suffers from notable decline between 3 and 8 months [21,22], suggesting that earlier survey intervals may be beneficial for capturing accurate response data.

Determining the accuracy of patient self-report is confounded by several additional factors. First, index institutions are only implicitly aware through their own reporting systems of readmissions back to their own health care systems. It has been reported, for example, that leakage—presentation of the patient to facilities other than the index facility for complications and readmissions—occurs in 31% to 65% of cases with some rates as high as 87.5%, suggesting that index institutions have large blind spots about postencounter health care utilization for which they may bear financial risk [16,17,23]. This degree of leakage, in a health care environment such as that of the United States which lacks a single payer, means that readmissions, complications, and health care utilization may be underrecognized and underreported. Although large public payers such as Medicare may be able to report readmission rates back to index facilities with reasonable leakage-free accuracy, Medicare beneficiaries are not demographically representative of the US population at large and constitute a portion of the US population that is increasingly being outpaced in certain procedural volume areas by other age cohorts [24]. Second, reliance solely on metrics such as proportion of correct reports [15], concordance [16,17,25], or agreement [13] that inflate when event prevalence rates are low rather than presenting these metrics alongside of an appropriate kappa statistic may lead to misinterpretations of the accuracy of patient self-report.

Some studies have attempted to address leakage by using single [17] or multi-institutional [25] databases or registries. Notable among them is Harrold et al [18], who used medical records from the Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) network, a group of over 230 surgeons across 28 states, to evaluate the accuracy of patient self-report following total knee and total hip replacement against data from hospitals within the region of the index facility as the reference. The study also evaluated the accuracy of patients reporting no utilization by examining orthopedic notes at the FORCE-TJR core sites as well as emergency department, day surgery, and hospitalization records at the index site. For 60% of the patients, the nearest

hospitals to their homes were not the index facilities, so the investigators received releases of information from the nearest hospitals to the patients' homes in 87% of cases. Nevertheless, utilization or complications documented in primary care or other specialty settings or in external facilities not otherwise included may have posed limitations.

Two parties, the payers and the patient, may be the most knowledgeable about health care utilization and complication events and in ideal positions to report the most comprehensive health care event information post encounter. If quality improvement is a goal, and readmissions and postencounter complications represent the last mile in the health care quality chain, could patients become active participants by providing accurate and timely information about these outcomes back to health care organizations? While payer databases may be nearly leakage-free references against which to compare patient self-report, relatively few have been used in such analyses, and when they have, they have been largely limited to single-payer settings [26-28] or to employer-based health care claims data [29] that do not necessarily generalize to the broader population.

To our knowledge, this is the first study of the accuracy of patient self-report using a commercial payer claims database as a reference that (1) minimizes the potential for underreporting due to leakage across specialties, care settings, and institutions, (2) enables measurement of self-report among patients attesting to either the presence or absence of events, (3) includes a nonresponse bias analysis, and (4) facilitates the timely collection of responses to mitigate recall error using workflow-compatible tools such as automated internet-based DPE platforms to survey patients in an optimal postencounter timeframe.

Methods

Study Design

We conducted a multicenter observational cohort study on postdischarge outcomes following orthopedic procedures falling into 1 of 5 categories: hip arthroplasty, knee arthroplasty, knee arthroscopic procedures, shoulder arthroscopic procedures, and knee arthrotomy. As part of a broader investigation of the impact of automated DPE platforms on health care costs and outcomes [30], data on patient self-report of 90-day hospital admissions, ERUC use, and complications were collected through an automated DPE platform (HealthLoop Inc) and compared against claims data from Anthem Inc for Anthem members who had undergone the procedure at 1 of 10 community orthopedic practices in California and Nevada between March 1, 2015, and July 1, 2016. These community practices ranged in size from solo practitioner to multispecialty practices with as many as 25 physicians.

We followed the Reporting of Studies Conducted Using Observational Routinely-Collected Data (RECORD) statement checklist (an extension of the Enhancing the Quality and Transparency of Health Research [EQUATOR] Network Strengthening the Reporting of Observational Studies in Epidemiology [STROBE] guidelines) [31,32] and the Standards for Quality Improvement and Reporting Excellence (SQUIRE) guidelines [33]. Although this was not a randomized controlled trial, we adhered to as many of the Consolidated Standards of Reporting Trials (CONSORT)-EHEALTH checklist items (v1.6.1) as appropriate [34,35]. This study received a determination of exemption from human subjects research by E&I Review Services, an independent institutional review board.

Enrollment

Practices whose patients were sent utilization and complication surveys were already using the DPE platform in their routine provision of care to provide asynchronous, automated remote guidance and conduct telemonitoring before and after procedures. Because this was a retrospective observational cohort study and not a prospective trial, there was no recruitment. Patients were enrolled for routine clinical purposes on the platform that was already in use at practice sites (ie, they were not enrolled in a trial), and investigators later compared deidentified survey responses to claims data. Enrollment of patients on the platform was at the discretion of the individual practices and was not within the influence of the authors. However, since practices were using the platform for their routine provision of care, most patients undergoing relevant procedures at these sites were enrolled on the platform and were receiving surveys. The only exclusion criteria at the points of care were the lack of a valid email address and internet access, as required to receive check-in notifications and interact with the DPE platform itself.

Digital Patient Engagement Platform and Health Care Utilization and Complication Surveys

For context, the DPE platform worked as follows. Automated email check-in notifications generated by the platform and designed to come from the physicians were sent to patients longitudinally over time according to predetermined procedure-specific care plan schedules. A notification link within the email prompt took the patient into the Health Insurance Portability and Accountability Act (HIPAA)-compliant environment where materials pertinent to that day, written to a Flesch-Kincaid 6th grade reading level, were queued up, including reminders, checklists, educational materials, structured symptom assessments, and patient reported outcome measure (PROM) surveys. At approximately 90 days post encounter, the utilization and complication questions pertinent to this study were asked of patients (see Multimedia Appendix 1).

With regard to health care utilization, patients were asked about hospital admissions and ERUC visits. They were also queried about complications relevant to their procedures, including deep vein thrombosis (DVT), pulmonary embolism (PE), surgical site infection (SSI, including sepsis), hemorrhage (including gastrointestinal bleeding), fracture/dislocation, and severe constipation. Delivery of surveys and collection of responses were fully automated, requiring no additional human support. Further DPE platform details have been described in Steele [36] and Rosner [30]. The platform was accessible to patients and health care professionals via the internet on desktop, laptop, tablet, and iOS- or Android-enabled mobile devices.

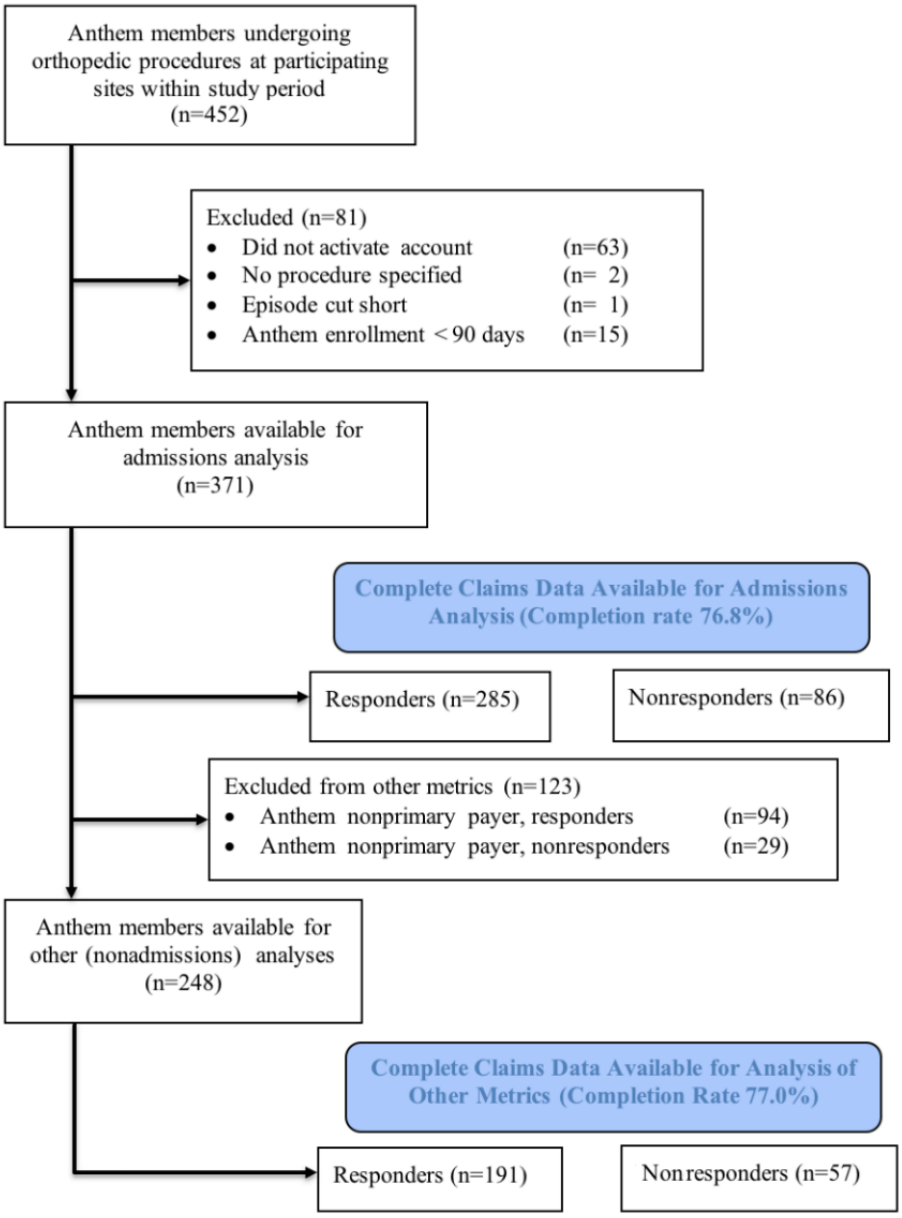
Data Inclusion and Exclusion

Claims outcomes were identified in the Anthem database with *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) and ICD-10-CM codes. Because the transition from ICD-9 to ICD-10 occurred during the study period, forward and reverse mappings between the versions were performed. All data extracted from the Anthem database were deidentified with randomized case identifiers applied for investigator reference.

Not considered for the study were practice patients who were not Anthem members, because we did not have claims data against which to compare patient responses. Among the Anthem members receiving surveys, excluded from analysis were patients not having a procedure within 1 of the 5 specified categories, patients whose enrollment with Anthem terminated prior to 90 days (ie, incomplete 90-day claims data), and patients with short episode durations, defined as patients who had more than 1 eligible procedure within a 90-day timeframe whose 90-day surveys for the first procedure could have overlapped with and been confounded by events associated with the second procedure (Figure 1). Response data from patients with capitated health maintenance organization products for which Anthem did not have full professional service claims were also to be excluded, but after the above exclusions, there were no remaining patients for whom this applied.

For admissions outcomes, the Anthem database contained complete data even for patients for whom Anthem was not the primary payer. However, for other outcomes (eg, ERUC visits and complications), the Anthem database did not necessarily contain full data for patients whose primary payer was not Anthem. Therefore, for outcomes other than admissions, data from patients for whom Anthem was not the primary payer were further excluded from analysis.

Figure 1. Data inclusion and exclusion waterfall.



Primary Analysis

The primary outcome metric for each survey question was the kappa statistic, a standard measure of how much the observed agreement between patient self-report and the events in the claims database differed from expected. Other metrics of relevance, widely reported in related studies, included true positive (TP), true negative (TN), false positive (FP), false negative (FN), sensitivity, specificity, positive predictive value, negative predictive value, and agreement, the latter of which was defined as (TP+TN)/(TP+TN+FP+FN). Although some studies use agreement as the primary outcome metric to evaluate the accuracy of patient self-report, agreement is prone to inflation when event prevalence rates are low, and kappa serves as a standard statistic that is not unduly influenced in this manner. Many studies have used the following thresholds as guidance to help interpret the meaning of the value of kappa, but there is not universal agreement as to what cut points should be considered clinically meaningful, and as such, interpretation of kappa in a relative sense is more useful than in an absolute sense: kappa <0.20, poor; 0.20 to 0.39, fair; 0.40 to 0.59, moderate; 0.60 to 0.79, very good; and ≥0.80, excellent [37-39]. A sensitivity analysis of the primary outcome (Figure 2) was also performed to illustrate the influence on kappa of changing 1 TN response to TP and 1 TN response to FP for each question.

Secondary Analyses

We conducted 3 secondary analyses to examine for potential bias that could influence kappa, the primary outcome metric. When it was not possible to evaluate for bias influencing kappa (eg, when there were no self-report data from which to calculate kappas from cohorts such as account nonactivators or survey nonresponders), we considered differences in surrogate metrics such as event prevalence rates or demographics. The 3 demographic variables to which the authors had access were age; gender; and DxCG score, a composite indicator of overall illness burden.

In the first among these, a nonactivator bias analysis, we examined for bias between patients who activated their platform accounts and those who did not. Since kappas were not available for comparison (nonactivators, by definition, did not furnish self-report data from which kappas could be calculated), we examined for differences in demographics between these cohorts. We did not examine for differences in event prevalence rates in this bias analysis since it has been shown in the literature [30] that one of the effects of DPE platforms for patients who activate their accounts is to reduce event rates relative to those who do not.

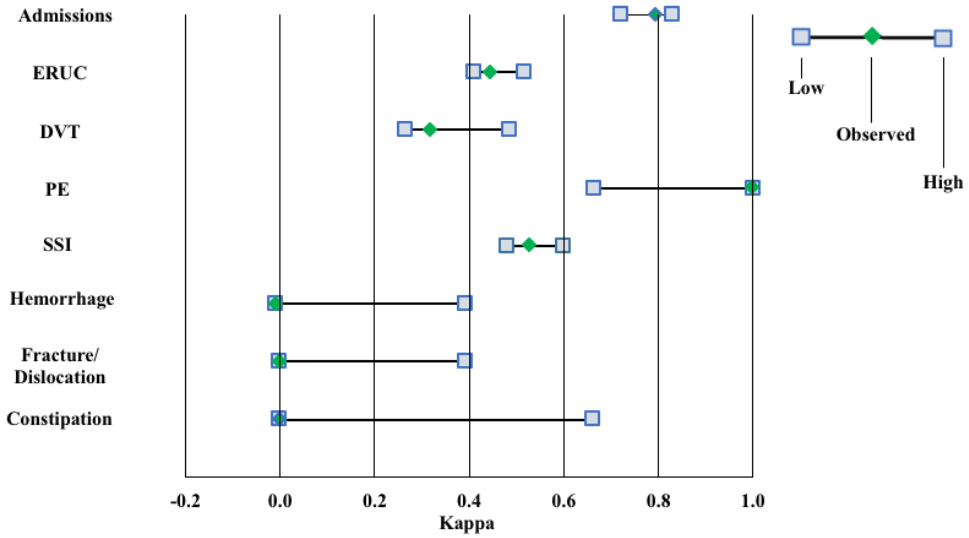
We similarly conducted a nonresponse bias analysis, examining for differences in demographics between activated patients who responded to self-report surveys and those who did not. Again, differences in kappa could not be examined because the nonresponders furnished no self-report data, but evaluation for differences in event prevalence rates was possible because account activation status in both cohorts was the same.

Finally, we conducted a bias analysis between 2 cohorts with expected demographic differences: the arthroplasty cohort (expected to be older) and the nonarthroplasty cohort. Since age and comorbidity burden are known to drive event rates, differences in event rates would not necessarily be an accurate assessment of bias. However, in this analysis, since both cohorts did furnish self-report data, we were able to directly assess for differences in kappa as a function of demographics.

Statistical Analysis

Analyses were performed in R version 3.2.3 (The R Foundation). Fisher exact test and analysis of variance were used where categorical and continuous variables were compared, respectively. $P<.05$ was deemed significant. Kappas were computed using the CohenKappa function from the R DescTools package and compared using the values and standard errors produced by that function. The kappa statistic was considered significant if the confidence interval excluded 0 [40].

Figure 2. Sensitivity of observed kappa to changing 1 true negative to false positive and to changing 1 true negative to true positive. DVT: deep vein thrombosis; ERUC: emergency room/urgent care; PE: pulmonary embolism; SSI: surgical site infection.



Results

User Statistics

The mean age of the 371 survey recipients available for admission analysis was 56.5 (SD 15.7) years, 48.8% (181/371) of whom were female. The mean DxCG score for this group was 5.32 (SD 5.28). The mean age of the 285 Anthem members who responded to the surveys was 56.9 (SD 15.4) years, 49.5% (141/285) of whom were female (Table 1). The mean DxCG score for the 285 responders was 4.89 (SD 4.96). As a measure of overall platform usage (not just survey response rates) within the responder cohort, the mean patient engagement, measured as the number of check-ins performed divided by the number of check-ins scheduled, with additional credit in both the numerator and denominator for proactive, unscheduled activity in the platform, was 79.7% (SD 19.9). There was no statistical difference in overall platform usage as measured by engagement between patients less than 65 years of age and those 65 years and older (*P*=.61).

Primary Analysis

Surveys were sent to 452 Anthem members, of whom 371 met the inclusion criteria (Figure 1). Of these 371 patients, 285 completed 1 or more survey questions and submitted the surveys regarding admissions (76.8% completion rate [41]). Regarding all other survey questions (for which availability of complete claims data required patients to have Anthem as a primary payer), 123 patients for whom Anthem was not the primary payer were excluded from analysis. Of these patients, 65.0% (80/123) had Medicare as their primary payer. Among the patients for whom Anthem was the primary payer, 248 met the inclusion criteria, and 191 completed 1 or more questions and submitted the survey (77.0% completion rate).

With regard to 90-day admissions, patient self-reports were found to be characterized by a kappa of 0.80 and agreement of 0.99 (Table 2). With respect to ERUC, patient responses were found to be characterized by a kappa of 0.45 and agreement of 0.96. Regarding complications, patient responses were characterized by kappas and agreements of 1.00 and 1.00 for PE, 0.53 and 0.97 for SSI, 0.32 and 0.97 for DVT, 0.00 and 0.98 for fracture/dislocation, 0.00 and 0.99 for severe constipation, and –0.01 and 0.98 for hemorrhage, respectively.

Secondary Analyses

In the nonactivator bias analysis, we found there were no significant demographic differences between the patients who activated their DPE platform accounts and those who did not. The mean ages in the activated versus nonactivated cohorts were 56.2 (SD 15.7) and 55.8 (SD 18.5) years, respectively (*P*=.88). The gender distributions in the activated versus nonactivated cohorts were 48.4% (188/388) female and 45.3% (29/64) female, respectively (*P*=.69). The mean DxCG scores in the activated versus nonactivated cohorts were 5.30 (SD 5.25) and 5.18 (SD 5.89), respectively (*P*=.87).

In the nonresponse bias analysis, there were no significant demographic differences between patients who responded and those who did not. The mean ages in the responder versus nonresponder cohorts were 56.5 (SD 15.4) years and 55.3 (SD 17.5) years, respectively (*P*=.46). The gender distributions in the responders versus nonresponders were 49.5% (149/301) female and 45.0% (68/151) female, respectively (*P*=.42). The mean DxCG scores in the responders versus nonresponders were 4.96 (SD 5.27) and 5.91 (SD 5.44), respectively (*P*=.08). The lowest *P* value for event prevalence rate differences between the 2 groups was .30 (Table 3).

Table 1. Demographic characteristics of responders.

Characteristic	Value
90-day admission (all payers, n=285)	
Age (years), mean (SD)	56.9 (15.4)
Age (years), median (IQR[a])	58 (47, 68)
Female, n (%)	141 (49.5)
DxCG[b], mean (SD)	4.89 (4.96)
DxCG, median (IQR)	3.58 (2.02, 6.25)
90-day events (Anthem primary payer, n=191)	
Age (years), mean (SD)	49.0 (12.3)
Age (years), median (IQR)	52 (42, 58)
Female, n (%)	92 (48.2)
DxCG, mean (SD)	3.99 (4.95)
DxCG, median (IQR)	2.87 (1.59, 4.77)

[a]IQR: interquartile range.

[b]DxCG: a composite indicator of overall illness burden.

Table 2. Counts and calculated values for determination of accuracy of patient self-report.

Characteristics	TP[a]	FP[b]	TN[c]	FN[d]	Prevalence	Sn[e]	Sp[f]	PPV[g]	NPV[h]	Agreement	Kappa (95% CI)
90-day admission (all payers)											
Admission	4	2	239	0	0.02	1.00	0.99	0.67	1.00	0.99	0.80 (0.52 to 1.00)[i]
90-day events (Anthem primary)											
Emergency room/urgent care visit	3	7	148	0	0.02	1.00	0.95	0.30	1.00	0.96	0.45 (0.11 to 0.78)[i]
Pulmonary embolism	1	0	120	0	0.01	1.00	1.00	1.00	1.00	1.00	1.00 (1.00 to 1.00)[i]
Surgical site infection	3	3	149	2	0.03	0.6	0.98	0.50	0.99	0.97	0.53 (0.17 to 0.89)[i]
Deep vein thrombosis	1	1	115	3	0.03	0.25	0.99	0.50	0.97	0.97	0.32 (–0.17 to 0.81)
Fracture/dislocation	0	3	152	0	0.00	N/A[j]	0.98	0.00	1.00	0.98	0.00 (0.00 to 0.00)
Severe constipation	0	1	154	0	0.00	N/A	0.99	0.00	1.00	0.99	0.00 (0.00 to 0.00)
Hemorrhage	0	2	154	1	0.01	0.00	0.99	0.00	0.99	0.98	–0.01 (–0.02 to 0.00)

[a]TP: true positive.
[b]FP: false positive.
[c]TN: true negative.
[d]FN: false negative.
[e]Sn: sensitivity.
[f]Sp: specificity.
[g]PPV: positive predictive value.
[h]NPV: negative predictive value.
[i]Indicates statistical significance.
[j]N/A: not applicable.

Table 3. Nonresponder bias analysis.

Characteristics	Responder (n)	Responder prevalence	Nonresponder (n)	Nonresponder prevalence	*P* value
90-day admissions (all payers)					
Admissions	245	0.02	40	0.00	>.99
90-day events (Anthem primary)					
Emergency room/urgent care visit	158	0.02	33	0.00	>.99
Pulmonary embolism	121	0.01	70	0.00	>.99
Surgical site infection	157	0.03	34	0.03	>.99
Deep vein thrombosis	120	0.03	71	0.00	.30
Fracture/dislocation	155	0.00	36	0.00	>.99
Severe constipation	155	0.00	36	0.00	>.99
Hemorrhage	157	0.01	34	0.00	>.99

In comparing arthroplasty to nonarthroplasty cohorts, we found no difference in gender distribution between the 2 groups (female 51.4% [53/103] vs 48.4% [88/182], respectively, P=.62). As expected, the arthroplasty patients were older (mean age of 66.7 (SD 10.2) years versus 51.3 (SD 15.0) years, respectively, P<.001) and had higher mean DxCG scores (7.36 [SD 5.70] versus 3.49 [SD 3.85], respectively, P<.001). However, despite the age and DxCG differences, there were no significant differences between these groups in kappa for any of the questions (lowest P value .09), suggesting that differences in age and illness burden across these cohorts did not have effects on self-report accuracy.

Discussion

Principal Findings

In this multicenter observational cohort study, we sought to assess the accuracy of patient self-report of health care utilization and complications in the 90 days post encounter following 5 types of orthopedic procedures. We found the accuracy of patient self-report of 90-day hospital admissions and 90-day ERUC visits to be characterized by kappas of 0.80 and 0.45, respectively. These findings are consistent with those of Ungar [42] (kappas of 0.80 and 0.60, respectively), who described parental report of pediatric asthma-related

hospitalizations and emergency room visits in a Canadian population, and Yu [28] (kappas of 0.75 and 0.52, respectively), who described self-report of utilization of such services among a general Taiwanese population.

We also found the accuracy of patient self-report of 90-day PE events, SSI, DVT, fracture/dislocation, severe constipation, and hemorrhage to be characterized by kappas of 1.00, 0.53, 0.32, 0.00, 0.00, and –0.01, respectively, although the interpretation of several of these items may be limited in our study by small sample sizes for events with extremely low prevalence rates. For example, the limitation related to fracture/dislocation is demonstrated, when sample size was larger in a New Zealand registry study [43], by the close agreement observed between patient self-report of hip dislocation and revision in the 6 months following hip arthroplasty and registry confirmed hip dislocation and revision (0.37% vs. 0.39%, respectively). This limitation is further demonstrated by a sensitivity analysis (Figure 2) in which we examine the kappa when 1 TN is changed to TP and when 1 TN is changed to FP. The kappa is shown to be particularly sensitive to events of lowest prevalence in which there are either no TPs or FPs in our sample (severe constipation, fracture/dislocation, hemorrhage, and PE). The implication, particularly for the first 3, is that due to low event prevalence, our results may not have sufficient resolution to conclude that patients are necessarily poor self-reporters of these specific complications.

It is noteworthy that accuracy appears quite high for some items and lower for others, even when event prevalence is not negligible. One explanation that has been suggested is the concordance between patient self-report and a reference is higher for significant events such as hospitalizations than for more routine events [29]. Our results may be consistent with that explanation. Nevertheless, it is noteworthy that 2 patients had false positive reports of admissions. One explanation is that while these patients may have been accurate in reporting an overnight stay in the hospital, for billing purposes they might have been classified as outpatients under observation status or under the Centers for Medicare and Medicaid Services' 2-midnight rule. To minimize the impact that clinical decision units or observation units might have on false positives, our survey question for hospital admissions (Multimedia Appendix 1) asked patients not to count overnight stays in the emergency department. Not all clinical decision units or observation units are physically located in emergency departments, however, and it may be difficult for patients to know or even later ascertain whether their stay in such a unit or their stay in the hospital for less than 2 midnights had been classified by the payer as inpatient or outpatient.

Regarding the lower accuracy of self-report observed for ERUC than for hospital admissions, the difference between payer classification and patient perception of what constitutes an urgent care visit may be central. It has been reported, for example, that patients may consider an urgent care visit to be either to an urgent care center or a general practitioner for a same-day visit. On the other hand, payer claims data differentiate the services based on location and would attribute the general practitioner visit as an outpatient encounter rather than an urgent care encounter [29]. Therefore, it is arguable that these 2 visit types should not be aggregated within a single survey question and that location should be more distinctly specified. This hypothesis seems to be supported by the high agreement and high kappas reported by Harrold et al [18] regarding emergency room visits post discharge in which the emergency room was the only location specified in the survey question.

Accuracy of self-report may also suffer due to survey question language around concepts that are well understood to medical practitioners but not to others. For example, Greenbaum [17] found that there was good concordance for clearly defined complications (eg, pulmonary embolism, dislocation) and poor concordance for less clearly defined complications (eg, major bleeding). Similarly, Bream [9] reported that accuracy of SSI self-reporting was variable, but there was greater accuracy when patients were asked about symptoms or antibiotic use (as we have done) rather than being asked about an overall diagnosis. This suggests that the limitations may not reside with the patient's actual capacity for accuracy but with the language and construct of the questions. Such language should be within the grasp and availability of the lay person, although for certain medical concepts, this may not be possible. When the language is put into lay terminology and in the context of phenomena within the patient experience, accuracy may be optimized.

Also of relevance are the intervals at which patients are asked to self-report. Although short intervals (eg, 30 days) might be desirable from a recall perspective, cumulative event rates at 30 days are likely to be low and less useful to health care organizations than event rates accumulated over longer periods of time. Furthermore, administering surveys at high frequencies and comparing them to references at recurring intervals such as 30, 60, and 90 days may be prohibitively resource intensive using traditional means. As such, many studies have asked patients for self-report at a single 6-month time point [16,17,44]. However, such a long lag between a health care utilization or complication event and the survey itself may introduce recall error. In several studies, stability of patient recall appears to remain over 2 to 3 months [19,20] but suffers from marked decline between 3 and 8 months [21,22]. Survey periods of 90 days, as in this study, may not only minimize recall error but facilitate accuracy and timeliness of results in closed loop feedback cycles to index institutions engaged in quality improvement.

Accuracy of patient self-report is just 1 component critical to postencounter quality improvement processes for health care organizations. Beyond accuracy are needs for easy distribution and collection of surveys, timely reporting to ensure a temporally proximate feedback cycle, and high rates of patient response. In this study, we report a 76.8% completion rate of 90-day surveys facilitated entirely through an automated process, requiring no additional manual support. Automated DPE platforms may offer a practical and scalable distribution and collection modality acceptable to patients and health care teams.

Strengths

This study has overcome several limitations of prior, related studies in that it (1) mitigated the potential for underreporting due to leakage across specialties, care settings, and institutions, (2) enabled the measurement of self-report among patients

attesting to the presence and absence of events—the latter being a major challenge in studies involving manual chart review because the cohort of patients without events is substantially larger than those with events—and (3) facilitated the timely collection of responses to mitigate recall error using workflow compatible tools.

Using a payer claims database overcame limitations inherent in a commonly applied technique of using physician retrospective report as a reference, an approach that has been described as prone to underreporting due to poor professional compliance with completing audit data, inaccurate coding of procedures [44], unawareness due to leakage [18,25], and potential for bias [45].

Another strength was the use of 90-day self-report surveys rather than longer periods commonly used such as 6 months [16,17,44], as it has been demonstrated that accuracy of self-report begins to taper after 2 to 3 months [20,21]. Finally, unlike many studies which fail to conduct nonresponse bias analyses—including those that acknowledge the potential of nonresponse bias in their own samples [15]—we did conduct such an analysis and found there to be no nonresponse bias in our sample.

Limitations

Accuracy of claims data is subject to the accuracy of coding, which is reportedly variable [46,47]. Low prevalence rates of some events in our sample also limit resolution of the results for several of the survey questions. The influence of demographic factors such as age, race, ethnicity, socioeconomic status, and level of education on the use of internet and email and on the potential for inaccuracy of self-report is worthwhile to consider. Several studies have demonstrated differences in the use of email and internet according to race and ethnicity [48] and based on age, with the most notable age-based use drop-off among those over 75 years [48,49]. The only demographic factors to which the authors had access, however, were age, gender, and DxCG score. Subgroup analysis based on age alone was not possible because the low prevalence rates of events in our sample combined with exclusion of data for patients whose primary payer was not Anthem (particularly those 75+ years of age for whom internet and email usage is reported to drop) rendered most subgroup analyses too small to lead to any meaningful conclusions. However, we did explore the influence of age and disease burden by comparing the older arthroplasty group to the younger nonarthroplasty group and found no difference in accuracy of self-report between these groups. An additional limitation was that comprehensibility of survey questions around concepts that are inherently clinical may be a factor in this sort of investigation. Although we made every attempt to put questions in terms within patients' grasp, some concepts will likely always be challenging for patients to self-report, either because the definition of an event is clinical or the patient does not have access to all of the information needed to self-report (eg, lab values, imaging studies, classification of a 1-night hospital stay as either inpatient or outpatient).

Generalizability is often a key issue in translating study findings to real-world practice. Our findings came from a limited set of orthopedic patients in a West Coast US geographic area and may not necessarily generalize to other patient populations, geographies, and medical conditions. However, there are other aspects of our study that may contribute toward generalizability. First, rather than being limited to a single site, this was a multicenter study that drew from community orthopedic practices. Second, while the lack of real-world practice results has often been criticized among digital health applications [50], this retrospective study occurred in real-world practice settings and did not involve recruitment, formal inclusion or exclusion criteria, or research staff to support or promote patient engagement.

Conclusions

We have demonstrated through the use of an internet-based automated DPE platform in real-world clinical settings kappa and agreement values for patient self-report of 90-day hospital admission of 0.80 and 0.99, respectively, and of 90-day ERUC visits of 0.45 and 0.96, respectively. We have also demonstrated higher accuracy for major complications such as PE and lower accuracy for complications such as hemorrhage, which were found to be subject to low event prevalence rates and small sample sizes. We further demonstrated a survey completion rate of 76.8%, requiring no additional support in real-world clinical practice settings, and that there was no significant bias introduced by platform nonactivators, survey nonresponders, or patients of older age or higher disease burden.

These findings may bear relevance to the very health care entities that are increasingly bearing risk under programs such as the Hospital Readmissions Reduction Program and bundled payment programs including the Comprehensive Care for Joint Replacement program, the Bundled Payments for Care Improvement initiative [1], and its recent successor, Bundled Payments for Care Improvement Advanced [2]. These institutions have had limited temporally proximate insight into readmission and postdischarge complication rates for their own patients, in part because of leakage and in part because of the lag between when an event happens and when reconciliation occurs. Patient self-report of utilization and complications has been considered not only as a means of enhancing the accuracy and timeliness of utilization and complication reporting but as a potential means of engaging the patient further as a partner in his or her own care.

As health care facilities consider such self-report mechanisms as means to enhance their own quality improvement efforts, capture of health care events by index institutions is only part of the needed solution. It remains up to these institutions to implement quality improvement initiatives that reduce potentially avoidable readmissions and complications based on the closed feedback cycle. Additional research spanning other medical specialties, geographies, and patient populations may demonstrate whether this approach could be generalized more broadly.

Acknowledgments

BIR designed the study, participated in the analysis, wrote the manuscript, and approved it for publication. MG queried claims databases and reviewed, contributed to, and approved the manuscript for publication. WNA conducted statistical analyses and reviewed, contributed to, and approved the manuscript for publication. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank Tom Bedor, Jannifer Harper, MD, Michael Kaplan, Tony Linares, MD, and Jordan Shlain, MD.

Conflicts of Interest

BIR is an employee of and reports equity interest in HealthLoop Inc. MG is an employee of and reports equity interest in Anthem Inc. WNA is a paid consultant of HealthLoop and reports equity interest in Amgen Inc and Edwards Lifesciences LLC.

Multimedia Appendix 1

Utilization and complications questionnaire.

[PDF File (Adobe PDF File), 21KB - jmir_v20i7e10405_app1.pdf]

References

1. Center for Medicare & Medicaid Innovation. Bundled Payments for Care Improvement (BPCI) initiative: general information URL: https://innovation.cms.gov/initiatives/bundled-payments/ [accessed 2018-06-28] [WebCite Cache ID 70WeUGnVY]
2. Center for Medicare & Medicaid Innovation. BPCI advanced URL: https://innovation.cms.gov/initiatives/bpci-advanced [accessed 2018-06-28] [WebCite Cache ID 70WeZhlvz]
3. Center for Medicare & Medicaid Innovation. Comprehensive care for joint replacement model URL: https://innovation.cms.gov/initiatives/cjr [accessed 2018-06-28] [WebCite Cache ID 70Wedz1bB]
4. National Compendium on Payment Reform. URL: http://compendium.catalyzepaymentreform.org/compendium-search/4585 [accessed 2018-03-13] [WebCite Cache ID 6xsaNhFAq]
5. Goozner M. Q&A with Dr. Patrick Conway URL: http://www.modernhealthcare.com/article/20180203/NEWS/180209999 [accessed 2018-03-14] [WebCite Cache ID 6xu5t4wi0]
6. McIlvennan CK, Eapen ZJ, Allen LA. Hospital readmissions reduction program. Circulation 2015 May 19;131(20):1796-1803 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.010270] [Medline: 25986448]
7. Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, observation, and the hospital readmissions reduction program. N Engl J Med 2016 Apr 21;374(16):1543-1551. [doi: 10.1056/NEJMsa1513024] [Medline: 26910198]
8. Centers for Medicare and Medicaid Services. Readmissions reduction program URL: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html [accessed 2018-06-28] [WebCite Cache ID 70WepslrK]
9. Bream E, Black N. What is the relationship between patients' and clinicians' reports of the outcomes of elective surgery? J Health Serv Res Policy 2009 Jul;14(3):174-182. [doi: 10.1258/jhsrp.2009.008115] [Medline: 19541877]
10. Rosner B. The hidden actors in hospital readmissions. Readmission News 2017;6(8):172.
11. St Sauver JL, Hagen PT, Cha SS, Bagniewski SM, Mandrekar JN, Curoe AM, et al. Agreement between patient reports of cardiovascular disease and patient medical records. Mayo Clin Proc 2005 Feb;80(2):203-210. [Medline: 15704775]
12. Tisnado DM, Adams JL, Liu H, Damberg CL, Chen W, Hu FA, et al. What is the concordance between the medical record and patient self-report as data sources for ambulatory care? Med Care 2006 Feb;44(2):132-140. [Medline: 16434912]
13. Liu B, Sweetland S, Beral V, Green J, Balkwill A, Casabonne D, Million Women Study Collaborators. Self-reported information on joint replacement and cholecystectomy agrees well with that in medical records. J Clin Epidemiol 2007 Nov;60(11):1190-1194. [doi: 10.1016/j.jclinepi.2007.02.007] [Medline: 17938062]
14. Parimi N, Lane NE, Bauer D, Hochberg MC, Nevitt MC. Accuracy of self-reported diagnosis of hip replacement. Arthritis Care Res (Hoboken) 2010 May;62(5):719-724 [FREE Full text] [doi: 10.1002/acr.20111] [Medline: 20191473]
15. Alazzawi S, Bardakos NV, Hadfield SG, Butt U, Beer ZH, Field RE. Patient-reported complications after elective joint replacement surgery: are they correct? J Bone Joint Surg Br 2012 Aug;94(8):1120-1125. [doi: 10.1302/0301-620X.94B8.29040] [Medline: 22844056]
16. Dushey CH, Bornstein LJ, Alexiades MM, Westrich GH. Short-term coagulation complications following total knee arthroplasty: a comparison of patient-reported and surgeon-verified complication rates. J Arthroplasty 2011 Dec;26(8):1338-1342. [doi: 10.1016/j.arth.2010.11.007] [Medline: 21256700]
17. Greenbaum JN, Bornstein LJ, Lyman S, Alexiades MM, Westrich GH. The validity of self-report as a technique for measuring short-term complications after total hip arthroplasty in a joint replacement registry. J Arthroplasty 2012 Aug;27(7):1310-1315. [doi: 10.1016/j.arth.2011.10.031] [Medline: 22197286]
18. Harrold L, Pascal S, Lewis C, O'Keefe R, Pellegrini V, Allison J, et al. Patient report improves posthospital discharge event capture in total joint replacement: a novel approach to capturing all posthospital event data. EGEMS (Wash DC) 2014;2(1):1107 [FREE Full text] [doi: 10.13063/2327-9214.1107] [Medline: 25848596]

19. Brown JB, Adams ME. Patients as reliable reporters of medical care process. Recall of ambulatory encounter events. Med Care 1992 May;30(5):400-411. [Medline: 1583918]
20. Carsjö K, Thorslund M, Wärneryd B. The validity of survey data on utilization of health and social services among the very old. J Gerontol 1994 May;49(3):S156-S164. [Medline: 8169350]
21. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. Med Care Res Rev 2006 Apr;63(2):217-235. [doi: 10.1177/1077558705285298] [Medline: 16595412]
22. Petrou S, Murray L, Cooper P, Davidson LL. The accuracy of self-reported healthcare resource utilization in health economic studies. Int J Technol Assess Health Care 2002;18(3):705-710. [Medline: 12391960]
23. Duseja R, Bardach NS, Lin GA, Yazdany J, Dean ML, Clay TH, et al. Revisit rates and associated costs after an emergency department encounter: a multistate analysis. Ann Intern Med 2015 Jun 02;162(11):750-756. [doi: 10.7326/M14-1616] [Medline: 26030633]
24. Wolford M, Palso K, Bercovitz A. National Center for Health Statistics. NCHS data brief, no. 186: Hospitalization for total hip replacement among inpatients aged 45 and over: United States, 2000-2010 URL: https://www.cdc.gov/nchs/products/databriefs/db186.htm [accessed 2018-03-14] [WebCite Cache ID 6xu6zecap]
25. Iyer R, Gentry-Maharaj A, Nordin A, Liston R, Burnell M, Das N, et al. Patient-reporting improves estimates of postoperative complication rates: a prospective cohort study in gynaecological oncology. Br J Cancer 2013 Aug 06;109(3):623-632 [FREE Full text] [doi: 10.1038/bjc.2013.366] [Medline: 23846170]
26. Lin C, Lai M, Syu C, Chang S, Tseng F. Accuracy of diabetes diagnosis in health insurance claims data in Taiwan. J Formos Med Assoc 2005 Mar;104(3):157-163. [Medline: 15818428]
27. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease. Comparing administrative data and self-reports. Med Care 1997 Sep;35(9):932-947. [Medline: 9298082]
28. Yu S, Chang H, Lin M, Lin Y. Agreement between self-reported and health insurance claims on utilization of health care: a population study. J Clin Epidemiol 2009 Dec;62(12):1316-1322. [doi: 10.1016/j.jclinepi.2009.01.016] [Medline: 19403264]
29. Palmer L, Johnston SS, Rousculp MD, Chu B, Nichol KL, Mahadevia PJ. Agreement between Internet-based self- and proxy-reported health care resource utilization and administrative health care claims. Value Health 2012 May;15(3):458-465 [FREE Full text] [doi: 10.1016/j.jval.2011.12.015] [Medline: 22583456]
30. Rosner BI, Gottlieb M, Anderson WN. Effectiveness of an automated digital remote guidance and telemonitoring platform on costs, readmissions, and complications after hip and knee arthroplasties. J Arthroplasty 2018 Apr;33(4):988-996 [FREE Full text] [doi: 10.1016/j.arth.2017.11.036] [Medline: 29229238]
31. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, RECORD Working Committee. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. PLoS Med 2015 Oct;12(10):e1001885 [FREE Full text] [doi: 10.1371/journal.pmed.1001885] [Medline: 26440803]
32. STROBE statement: available checklists. URL: https://www.strobe-statement.org/index.php?id=available-checklists [accessed 2018-03-14] [WebCite Cache ID 6xu7VM8NK]
33. SQUIRE 2: revised standards for quality improvement reporting excellence. URL: http://squire-statement.org/index.cfm?fuseaction=Page.ViewPage&PageID=471 [WebCite Cache ID 6xu8eth0t]
34. CONSORT-EHEALTH checklist v1.6.1. URL: https://asset.jmir.pub/assets/public/CONSORT-EHEALTH-v1-6.pdf [accessed 2018-07-11] [WebCite Cache ID 711NB30A0]
35. Baker TB, Gustafson DH, Shaw B, Hawkins R, Pingree S, Roberts L, et al. Relevance of CONSORT reporting criteria for research on eHealth interventions. Patient Educ Couns 2010 Dec;81 Suppl:S77-S86 [FREE Full text] [doi: 10.1016/j.pec.2010.07.040] [Medline: 20843621]
36. Steele JR, Jones AK, Clarke RK, Shiao SJ, Wei W, Shoemaker S, et al. Use of an online education platform to enhance patients' knowledge about radiation in diagnostic imaging. J Am Coll Radiol 2017 Mar;14(3):386-392. [doi: 10.1016/j.jacr.2016.11.018] [Medline: 28111051]
37. Bozic KJ, Chiu VW, Takemoto SK, Greenbaum JN, Smith TM, Jerabek SA, et al. The validity of using administrative claims data in total joint arthroplasty outcomes research. J Arthroplasty 2010 Sep;25(6 Suppl):58-61. [doi: 10.1016/j.arth.2010.04.006] [Medline: 20570479]
38. Stirnemann PM, Takemoto SK, Schnitzler MA, Brennan DC, Abbott KC, Salvalaggio P, et al. Agreement of immunosuppression regimens described in Medicare pharmacy claims with the Organ Procurement and Transplantation Network survey. J Am Soc Nephrol 2006 Aug;17(8):2299-2306 [FREE Full text] [doi: 10.1681/ASN.2006030258] [Medline: 16825328]
39. Altman D. Practical Statistics for Medical Research. London: CRC Press; 1990.
40. Fleiss J, Levin B, Paik M. Statistical Methods for Rates and Proportions. Hoboken: Wiley, John & Sons, Incorporated; 2003.
41. Eysenbach G. Improving the quality of Web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
42. Ungar WJ, Davidson-Grimwood SR, Cousins M. Parents were accurate proxy reporters of urgent pediatric asthma health services: a retrospective agreement analysis. J Clin Epidemiol 2007 Nov;60(11):1176-1183 [FREE Full text] [doi: 10.1016/j.jclinepi.2007.02.002] [Medline: 17938060]

43. National Joint Registry: 7-year report.: National Joint Registry, New Zealand Orthopaedic Association; 2007. URL: https://nzoa.org.nz/system/files/NJR%207%20Year%20Report.pdf [accessed 2018-07-11] [WebCite Cache ID 6xu9FMXAW]
44. Cockbain A, Carolan M, Berridge D, Toogood G. Performance and quality indicators: the importance of accurate coding. Bull R Coll Surg Engl 2012;94(2):46-50. [doi: 10.1308/147363512X13189526439151]
45. Khanna G, Singh JA, Pomeroy DL, Gioe TJ. Comparison of patient-reported and clinician-assessed outcomes following total knee arthroplasty. J Bone Joint Surg Am 2011 Oct 19;93(20):e117. [doi: 10.2106/JBJS.J.00850] [Medline: 22012534]
46. White RH, Garcia M, Sadeghi B, Tancredi DJ, Zrelak P, Cuny J, et al. Evaluation of the predictive value of ICD-9-CM coded administrative data for venous thromboembolism in the United States. Thromb Res 2010 Jul;126(1):61-67. [doi: 10.1016/j.thromres.2010.03.009] [Medline: 20430419]
47. Mears SC, Bawa M, Pietryak P, Jones LC, Rajadhyaksha AD, Hungerford DS, et al. Coding of diagnoses, comorbidities, and complications of total hip arthroplasty. Clin Orthop Relat Res 2002 Sep(402):164-170. [Medline: 12218480]
48. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
49. Smith A. Older adults and technology use. Washington: Pew Internet and American Life Project; 2014. URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use [accessed 2018-07-11] [WebCite Cache ID 6ytvT8Ysy]
50. Kvedar JC, Fogel AL. Why real-world results are so challenging for digital health. 2017. URL: https://catalyst.nejm.org/real-world-results-digital-health-products/ [accessed 2018-06-18] [WebCite Cache ID 70GFM3qYu]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DPE: digital patient engagement
DVT: deep vein thrombosis
DxCG: a composite indicator of overall illness burden
EQUATOR: Enhancing the Quality and Transparency of Health Research
ERUC: emergency room/urgent care
FN: false negative
FORCE-TJR: Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement
FP: false positive
HIPAA: Health Insurance Portability and Accountability Act
IQR: interquartile range
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
PE: pulmonary embolism
PROM: patient reported outcome measure
RECORD: Reporting of Studies Conducted Using Observational Routinely-Collected Data
SQUIRE: Standards for Quality Improvement and Reporting Excellence
SSI: surgical site infection
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TN: true negative
TP: true positive

Edited by G Eysenbach; submitted 18.03.18; peer-reviewed by J Walsh, J Black; comments to author 09.04.18; revised version received 23.04.18; accepted 21.06.18; published 20.07.18.

Please cite as:
Rosner BI, Gottlieb M, Anderson WN
Accuracy of Internet-Based Patient Self-Report of Postdischarge Health Care Utilization and Complications Following Orthopedic Procedures: Observational Cohort Study
J Med Internet Res 2018;20(7):e10405
URL: http://www.jmir.org/2018/7/e10405/
doi:10.2196/10405
PMID:30030212

Original Paper

Examining the Complexity of Patient-Outpatient Care Team Secure Message Communication: Qualitative Analysis

Holly Jordan Lanham[1,2,3], MBA, PhD; Luci K Leykum[1,2,3], MSc, MBA, MD; Jacqueline A Pugh[1,3], MD

[1]Division of General and Hospital Medicine, Department of Medicine, University of Texas Health San Antonio, San Antonio, TX, United States

[2]Department of Information, Risk and Operations Management, McCombs School of Business, The University of Texas at Austin, Austin, TX, United States

[3]Veterans Evidence Based Research, Dissemination, and Implementation Center (VERDICT), South Texas Veterans Health Care System, San Antonio, TX, United States

Corresponding Author:
Holly Jordan Lanham, MBA, PhD
Division of General and Hospital Medicine
Department of Medicine
University of Texas Health San Antonio
7703 Floyd Curl Drive
San Antonio, TX, 78229
United States
Phone: 1 210 567 7000
Email: lanham@uthscsa.edu

Abstract

Background: The value of secure messaging in streamlining routine patient care activities is generally agreed upon. However, the differences in how patients use secure messaging, including for communicating both routine and nonroutine issues, and the implications of these differences in use are less well understood.

Objective: The purpose of this study was to examine secure messaging use to extend current knowledge of how this tool is being used in outpatient care settings and generate new research questions to improve our understanding of the role of secure messaging in the patient-provider communication toolbox.

Methods: We conducted an in-depth qualitative analysis of secure message threads in 12 US Department of Veterans Affairs outpatient clinics in south Texas. We analyzed 70 secure message threads with a total of 179 unique communications between patients and their outpatient teams for patterns in communication and secure message content. We used theories from information systems and complexity science in organizations to explain our observations.

Results: Analysis identified content relating to 3 main themes: (1) information management, (2) uncertainty management, and (3) patient safety and engagement risks and opportunities. Within these themes, we identified 2 subcategories of information management (information exchange and problem solving), 2 subcategories of uncertainty management (relationship building and sensemaking), and 3 subcategories of patient safety and engagement risks and opportunities (unresolved issues, tone mismatch, and urgent medical issues). Secure messages were most often used to communicate routine issues (eg, information exchange and problem solving). However, the presence of subcategories pertaining to nonroutine issues (eg, relationship building, sensemaking, tone mismatch, urgent issues, and unresolved issues) requires attention, particularly for improving opportunities in outpatient care settings using secure messaging.

Conclusions: Patients use secure messaging for both routine and nonroutine purposes. Our analysis sheds light on potentially new patient safety concerns, particularly when using secure messaging to address some of the more complex issues patients are communicating with providers. Secure messaging is an asynchronous communication information system operated by patients and providers who are often characterized as having significant differences in knowledge, experience and expectations. As such, justification for its use beyond routine purposes is limited—yet this occurs, presenting a multifaceted dilemma for health care organizations. Secure messaging use in outpatient care settings may be more nuanced, and thus more challenging to understand and manage than previously recognized. New information system designs that acknowledge the use of secure messaging for nonroutine and complex health topics are needed.

(J Med Internet Res 2018;20(7):e218) doi:10.2196/jmir.9269

KEYWORDS

secure messaging; patient-physician communication; complexity science; outpatient care; outpatients; confidentiality

Introduction

Background

Secure messaging is rapidly becoming a commonly used health information technology (IT) tool [1]. This electronic communication feature, embedded within a patient portal, allows patients to communicate privately and securely with members of their care team about their health and medical condition(s), as well as about administrative matters such as medication refills and appointment requests. Despite reports of provider apprehension that use of secure messaging would increase workload [2-6], both patients and providers increasingly regard this health IT tool as an effective way to streamline health care delivery [7-10]. Patients who use secure messaging report higher patient satisfaction, improved face-to-face visits, and improved access to care outside of traditional in-person clinical visits [11]. Providers report positive impacts of secure messaging as well, particularly in terms of streamlining medication refills, managing referral requests, and scheduling appointments [4,12].

Features of Secure Messaging

Recent research has begun to highlight some important strengths of secure messaging tools for disease management, including a study showing that the use of secure messaging for prescription refills was associated with greater control of HIV viral load [13]. At the same time, studies identified key factors contributing to patient-level differences in secure message use, including end-user goals, internet availability, health literacy, and computer literacy [14]. At an organizational level, human resources, technology resources, and leadership support are associated with increased secure message adoption rates; higher secure message use is associated with lower urgent care use; and early adopters of secure messaging experienced a greater decrease in urgent care use over time than did later adopters [15]. Another study pointed to the perplexing nature of the IT-supported patient-provider relationship, finding that patients were responsive to provider engagement with secure messaging. Patients were more likely to use secure messaging if their provider frequently initiated messages to patients in general. If a provider was a low initiator, their patients were likely to be infrequent users of secure messaging as well [16].

Evidence of the value of secure messaging in streamlining routine patient care activities is growing [17-19]. While a substantial portion of the secure messaging literature has focused on describing the types of activities for its use, it has placed little emphasis on examining the complexity of these various activities and how different levels of message complexity might affect communication between patients and their providers. Secure messaging, because it is an asynchronous and virtual communication channel, is a lean form of communication lacking the capacity to convey the typical cues that characterize interpersonal conversation [20,21]. Gestures and nonverbal nuances, cues of social influence, symbolic content, and contextual cues are not easily captured and transmittable in secure messages. Thus, effective communication in an asynchronous technology-mediated context requires a great deal of effort and attention [22,23].

The effective transfer of rich information and the communication of ambiguous information via secure messaging is not well understood. Likewise, the linkage between secure message use and patient safety, and between secure message use and patient engagement, is not well understood. While researchers have examined secure message use in terms of message volume, frequency, and response time [12] and described activities for which secure messaging is used (eg, medication refills, appointment scheduling, referral request, and questions about medical conditions) [3,24,25], less is known about the nature of the information being exchanged and sought, and about capabilities of secure messaging for conveying information that is complex or nonroutine. We focused our examination on this aspect of secure messaging in an effort to extend current knowledge of how this tool can support outpatient care delivery, particularly for understanding the potential patient safety and engagement implications of using secure messaging to address nonroutine tasks and complex issues.

Conceptual Framework

We used complexity science [26-29] to frame our analysis because of its emphases on examining the interdependencies between system elements [30-32] and uncertainty management [31,33]. Complexity science is a useful lens through which to study relationships among primary care providers [34,35], care improvement in nursing homes [36], and collaboration in intensive care units [37]. In addition to its application in studying a diversity of health care delivery settings, complexity science has been useful in examining provider-level differences in health IT use [31], examining clinic-level differences in the implementation of health IT for practice redesign [38], and developing a sociotechnical model for studying health IT in health care systems [32]. Complexity science helped us evaluate secure messaging interactions between patients and their outpatient care team as a system of relationships [39] sometimes characterized by high complexity (contexts that cannot be fully understood simply by analyzing individual components of the system) [27] and uncertainty (an inevitable and natural part of complex systems that cannot be avoided, eliminated, or controlled) [40,41]. Thus, our study was both theoretically driven and grounded in the reality of everyday technology-supported communication between patients and their care teams.

Methods

Study Design

We conducted a retrospective in-depth qualitative analysis of secure message threads sent between patients and provider teams in 12 US Department of Veterans Affairs (VA) outpatient clinics of a single VA health system in south Texas, USA. Study clinics consisted of 9 primary care clinics, 1 mental health clinic, 1 allergy clinic, and 1 geriatric evaluation and management clinic.

The VA is undergoing a series of transformational initiatives to design a veteran-centric health care model and infrastructure to help veterans navigate the health care delivery system and receive coordinated care [42]. Health IT is a major part of this transformation, including the rollout of secure messaging as an additional communication channel for patients and their providers. This particular VA system uses a triage team model in its implementation of secure messaging: patients contact their care team and a nurse triages the messages, escalating messages to physicians as needed. We used a combination of thematic content analysis [43] and medical record audits. To follow up on any issues that appeared unresolved in the secure message thread, we examined patient medical records for signs of patient-provider communication that occurred through modalities other than secure messaging. The University of Texas Health Science Center San Antonio and South Texas Veterans Health Care System, San Antonio, TX, institutional review boards approved this study.

Data Collection

From the 12 study clinics, we retrospectively collected all secure messages sent between May 19, 2013 and December 19, 2013. We selected this time period to allow us to collect and analyze an adequate number of messages. At the time of the study, secure messaging was in the early phases of rollout and implementation at this VA. The only inclusion criterion was that the message was initiated by a patient of 1 of the 12 clinics during this 7-month time period. Data collection, achieved via a query sent to a clinical systems analyst followed by manual review, resulted in 70 total message threads with 179 unique messages and between 5 and 8 message threads per clinic. Each secure message thread was initiated by a unique patient and contained between 1 and 7 unique messages between a patient and their care team. The messages were captured in Word, deidentified and printed out for analysis. We excluded no messages in the analysis.

Analysis

We qualitatively analyzed [44] messages in 3 phases: (1) content analysis focused on uncovering general themes in the data and then developing subcategories under each theme, (2) systematic coding of the data, and (3) medical record auditing.

Content Analysis

The first phase was a content analysis using an open coding approach [45]. Two researchers (HJL, JAP) read and coded all messages. Messages were read to identify patterns in the types of information being exchanged or sought (requested). We abstracted text segments into a coding matrix to help with data sorting. During analysis, all 3 authors met to review and reach agreement on selected segments and the codes.

As themes emerged, we used complexity science literature in information systems and organizational sciences to interpret and refine themes and subcategories. For instance, medication renewals and appointment scheduling are message types that have been covered in the literature by previous studies of secure messaging, as information exchange and problem solving are known *information management* activities [46,47]. As such, we used this literature to define and examine these categories. We defined *information exchange* as content that is primarily aimed at sharing or transferring information between parties. We defined *problem solving* as content that presents a problem to be addressed.

We also identified patterns in the use of secure messaging that we coded as relationship building and sensemaking. For instance, we observed patients and providers who seemed to be using secure messaging as a way to establish or maintain the patient-provider relationship (relationship building). Further, we observed content in the messages where patients expressed confusion about their medical situation and sought help in interpreting or assigning meaning to something they were experiencing or to information they discovered from the patient portal or from another information source (sensemaking). Relationship building and sensemaking are known strategies for *uncertainty management* [33,40,48,49]. We used complexity science literature in organizational science and information systems [50,51] to define and examine these categories. We defined *relationship building* as content that sought to establish or maintain a relationship between parties. We defined *sensemaking* as content that demonstrated the seeking of new understanding or meaning, or help with interpretation of complex or ambiguous information.

Finally, we identified themes in the data pertaining to potential *patient safety* and *patient engagement* risks (or opportunities for improvement). For instance, we observed delays in care team responses to patient-initiated messages, tone mismatches between patient and care team messages, and urgent issues being communicated by patients via secure message. We viewed such content as having the potential to introduce unanticipated safety risks and detrimental effects on patient engagement. Therefore, we also coded the messages for these 3 safety and patient engagement subcategories. We defined *unresolved problems* as problems initiated in a secure message thread that were sometimes not resolved in that same thread and may have gone unresolved. We defined *tone mismatch* as messages from patients that included personal, emotional, or mental health details, or that were of a style that provided abundant detail; care team responses to these messages were brief or curt in tone, in contrast to the tone or content of patient-initiated messages. We defined *urgent medical issues* as messages that contained text with urgent or highly complex medical issues needing immediate medical attention.

Systematic Coding

In the second phase of analysis, we systematically coded the secure message threads for the 2 information management subcategories (information exchange, problem solving), 2 uncertainty management subcategories (relationship building, sensemaking), and 3 patient safety and patient engagement subcategories (unresolved issues, tone mismatch, urgent medical issues) that we had identified in the first phase of analysis [44]. Two authors (HJL, JAP) independently coded the messages and a third author (LKL) provided an additional perspective to resolve coding discrepancies and reach conceptual agreement. We discussed coding in 5 group sessions to ensure adequate consistency in the application of the coding definitions. Discussion with all 3 authors led to the final coding scheme,

and all authors reviewed the final coding for consistency and accuracy. Because threads frequently contained multiple messages, some threads were coded for multiple categories. Longer and more complex unique messages were often coded for multiple categories.

Medical Record Auditing

In the third phase of analysis, 2 authors (JAP, HJL) conducted medical record audits to follow up on messages that appeared unresolved after analyzing the secure message threads and messages categorized as urgent. The goal was to determine whether the issue or issues raised via secure message was or were ultimately addressed outside of the original secure message thread (eg, via an office visit, scheduled subspecialty outpatient visit, phone call). To do this, we searched for and reviewed additional secure messages sent after the data collection period to see whether the issue in question was resolved in a subsequent message. We also searched the medical record for follow-up appointments and office visits related to the issue raised via secure message. For example, if a patient asked for a referral to a physical therapist and the issue was unresolved in the original thread, we looked for a physical therapy visit or appointment scheduled close to the original request made via a secure message. We examined the medical record for anything that would signal or provide data that the issue was ultimately resolved. In this step, we considered all subsequent secure messages, visits, consults, and phone calls within 3 months of the initial secure message to be potentially involved in resolving an issue that appeared to have been unresolved in the original secure message thread.

Results

We identified and analyzed patient-outpatient care team secure message communication for 7 subcategories of secure messaging use nested within 3 main themes. We identified 2 information management subcategories, 2 uncertainty management subcategories, and 3 patient safety and engagement subcategories. Table 1 provides example quotes from the messages and the total number of threads coded for each category. Each quote was obtained from a unique patient. Approximately 50% (6/11) of the messages initially categorized as unresolved remained unresolved 3 months following the initial secure message communication.

Information and Uncertainty Management Strategies

We categorized secure message content as information management (information exchange and problem solving) and uncertainty management (relationship building and sensemaking). Information exchange (37/70, 53%) and problem solving (29/70, 41%) were more prominent in the data than sensemaking (10/70, 14%) and relationship building (6/70 instances, 8.6%).

Patient Safety and Patient Engagement Risks and Opportunities

We categorized secure message content related to patient safety and engagement as resolved or unresolved, matched or unmatched in tone between the patient's secure message content and the outpatient care team's secure message content, and urgent or nonurgent. We observed instances (11 out of 70) where issues raised over secure message appeared unresolved. It was not possible to tell from the original thread whether the problem was resolved in another thread, for example, or by a phone call or face-to-face visit—or if the problem truly went unresolved.

We observed tone mismatches between patients and outpatient care teams in 11 out of 70 messages. This most often occurred when a patient provided rich or personal details in their message to the care team and the care team responded using a template-type response, such as "Noted, will forward this to your provider." Other times, care team responses to highly emotional messages from patients were brief and curt in tone.

A less frequent type of message (3 out of 70) contained urgent medical questions from patients to their outpatient care teams (eg, seriously out of control blood pressure, suicidal thoughts). This observation suggests the need for further examination of the circumstances under which patients decide to use secure messaging for urgent medical matters, a growing patient safety concern [52].

Medical record reviews found that approximately 50% (6/11) of the messages categorized as unresolved remained unresolved after the medical record review. Issues that were resolved were addressed by phone or in-person visit. Of the 3 urgent secure messages, we also categorized 1 as unresolved and it appeared to remain unresolved after the medical record audit. The other 2 urgent messages were resolved by phone.

The findings generated from the medical record review supplemented the findings from coding analysis. Additionally, this step provided further insight into potential patient safety risks involved in unanticipated secure messaging use by patients.

Table 1. Secure message use subcategories and exemplar quotes in 70 message threads.

Subcategory	n (%)	Example content
Information exchange	37 (53)	*...would like to request a consult to be placed with physical therapy for "dolphin stem" treatment to help with scar tissue buildup post total knee replacement. Thank you.*
		I have tried to call and cancel throughout the weekend with no avail. I will not be able to make my appointment this afternoon. Please be sure to cancel it for me. Thank you.
Problem solving	29 (41)	*I have a nasty head cold. My nose is running constantly, sneezing, ache all over. Is there any over the counter meds I can take to help that won't react with the medications I'm taking?*
		...my omeperzole [sic*] has change there [* sic*] giving only filled my last script with 1 cap 10 mg a day. I need my coverage [3 times a day] due to my frequent feeding cause of my gastric bypass surgery. I have heart burn without why was it changed. Also can you send prosthetic a script for diabetic shoes they said my last script expired. Thank you.*
Relationship building	6 (8.6)	*I would like to set up this line of communication so that my appoints in the future will not be overlooked. Also I would like to apologize to _____ staff for my forgetting and missing my 11/19/13 appointment. Now that I have access to this Web page all my important information is in one place. Sincerely, _____*
		Hey there young man. You all ready for Christmas? If you are, you got me beat. _____, you didn't do a dam thing wrong my friend. Something was blocking the messages from coming through to me, that's it. Now as you can tell, everything is back to working just fine. Thank you for your help and patience.
Sensemaking	10 (14)	*Hi Team _____, I had a [computed tomographic] scan of my chest last week and was able to look at the results online.* Saw some words that make me uneasy, can you give me a quick email with your impressions and summary?
		Dr _____, I just wanted you to know that I had my Methacholine challenge test yesterday. I was confused when *the tech said it showed I DON'T have asthma. I was wheezing and a 72-year-old lifetime smoker by the last test. Then she gave me a dose of Albuterol, which cleared me right up and enabled me to blow the last spirometry test away. It that tightness and wheezing was not asthma, then what was it? I know we'll be able to talk about this next week at our appointment.*
		The reason I kept going to my mental health doctor, was not because I wanted to, it was because I needed to. I have serious problems with depression. I cry for no reason and have thoughts of suicide, I just want to lay in bed and do nothing, and I don't even want my son (who I love with all my heart) around me. I am taking Fluoxetine on a regular basis now and I'm still having bouts of depression. I really wasn't relaying this very well with my doctor, mainly because I wasn't having a "dark day" when I saw him. I need something to help with all these bouts that I have. It's an ongoing thing. Please help. I don't know why I keep having these.
Unresolved problems	11 (16)	*I have been seeing double vision 3 or 4 times every day for 2 to 4 minutes each time. For the past week I have been getting light headed just doing chores. My carpal tunnel supports need replacement please both of them. Thank you.*
		Response: none
Mismatches in tone	11 (16)	Long, detailed, multiproblem message with short response from patient-aligned care team nurse: "Will forward your concern to the doctor."
Urgent medical issues	3 (4.3)	*Dr ____, This morning I was to have an endoscopy but it was cancelled due extremely high blood pressure. I am faithfully taking my meds each morning around 9-9:30. I took the pills as directed this morning at 6am and arrived at the VA around 6:45am. My blood pressure was 208/110 and came down to 186/100 and then back up to the 200+ range. The endoscopy was cancelled. The chief of endoscopy was quite concerned as I was because I took my meds and have been taking them like I said – every morning. Now, I have had a lot of stress in the last 3 weeks. My father died and my brother and I are trying to get things…*
		Patient then goes on to describe death in the family plus other somber matters.

Discussion

Principal Findings

In this qualitative analysis of secure message communication between patients and providers in 12 outpatient clinics, we observed patterns in message content relating to secure message type and purpose beyond previously reported barriers and facilitators of secure message use, impacts on clinical workflow, and impacts on efficiency. We found secure message content to be straightforward and unambiguous in most messages. Patients used secure messaging as one might expect: for example, to check the status of a laboratory test result or to request a medication renewal. However, many messages were complex and multipurposed, often containing nonmedical, personal, or contextual details about the patient's life and social or personal situation. Others contained ambiguous or more complicated, less routine medical content that may not be easily addressable with a lean communication tool such as secure messaging [20].

Our analysis generated new questions about the use of secure messaging for nonroutine health care tasks and about how patients and care teams use secure messaging to communicate more complex and ambiguous information. Some patients shared highly personal and emotional content in their messages, others expressed discomfort with uncertainty in their medical condition, and a few patients conveyed urgent medical matters to their

outpatient care teams via secure messaging despite being advised against it. Provider concerns about these issues are not new. Studies describe provider concerns about messages that may be long, vague, and difficult to answer or inappropriately urgent [53-58]. Our study, however, provided evidence that such provider concerns are valid. This finding is important, as the tendency may be to accuse providers of expressing "concern" as a way to avoid using secure messaging tools for communicating with patients. Are secure messaging platforms set up to manage the exchange of information relating to more complex, nonroutine issues? If not, what interventions can be made from an organizational- or policy-level perspective to improve the ability of secure messaging platforms to communicate this type of information or to help manage the risk to patients' health if urgent matters are inappropriately communicated via secure messaging?

We also found tone mismatch and unresolved issues in our data, showing these lapses as examples of ineffective secure message patient-provider communication. Our findings generated new knowledge about the content of secure message conversations between patients and their providers and suggest potential links between secure message use and patient safety and patient engagement.

Sensemaking and Secure Messaging

Sensemaking is the process of assigning meaning to an unexpected event [42,44]. We observed clear examples of patients trying to make sense of their medical condition through secure message communication. Sensemaking may become problematic in cases where patients believe they are messaging their provider but are actually messaging a triage nurse. The disconnect between who the patient thinks they are messaging and the person who actually reads the message may contribute to mismatches in tone and then to unintended negative impacts on patient engagement. Managing this divide may be difficult, and it will likely depend on the delivery system. Nonetheless, patients need to know ahead of time with whom they are communicating when they engage in secure messaging because the physical and verbal cues present in face-to-face visits, telehealth technologies, and telephone messaging are absent.

The act of a nurse escalating a message to the provider holds clues for us about sensemaking and how to manage it in the context of secure messaging. Better understanding of what triggers the escalation of a message could add to our understanding of how outpatient care teams work together to develop a shared mental model of their patients and the actions needed to help patients be healthy. Similarly, we need better information management tools and policies for helping nurses and physicians respond to messages where a patient is expressing uncertainty or struggling to make sense of their medical situation. Likewise, knowledge is needed of when patients are using secure messaging as a tool to understand their own medical situation versus when patients are trying to connect with their care team so they can collectively make sense of the situation. A potential barrier to improving secure message use for sensemaking purposes is that dealing with messages containing this type of content is unlikely to save system time, just as playing phone tag for days decreases efficiency. We believe it likely that these are the sensemaking-oriented messages that nurses and physicians complain about when they express negative perceptions of communicating with patients by secure message. Given the nature of these messages, a richer communication channel, such as face-to-face or synchronous communication, is better suited. Another model might be to have a secure message in which sensemaking content is detected trigger a nurse message requesting a time to talk with the patient by phone. Regardless, we need better understanding of why patients use secure messaging to communicate complex and ambiguous information and IT communication tools designed to help providers manage this type of information from patients.

Tone Mismatch and Patient Engagement

Our analysis highlighted concern about mismatch in tone between the messages written by patients and the responses written by their care team. Patient messages in our dataset were generally received and triaged by a nonphysician, and the patients may have been unaware that their physicians were not actually the first people to receive their messages. This potential disconnect may have been a factor in the messages that were tone mismatched. It is also possible that the individuals responsible for triaging patient messages may not have understood the importance of their role in establishing and maintaining rapport with patients over secure message, or that because they are working from a computer (sometimes for long stretches of time) they temporarily forget that they are communicating with a patient who needs their help. We often think of secure messaging as a way to increase patient satisfaction, but if the response patients receive is uninviting or unconcerned, patient satisfaction may decrease. Repeated exposures to tone mismatch could result in patients refusing secure messaging tools, thus creating long-term challenges for organizations wanting to use this tool to communicate with patients.

Urgent Issues and Patient Safety

Despite a small number of urgent issues raised by patients in our dataset, they did exist. Of our 70 secure message threads, 3 contained an urgent medical matter. This number, while small, demonstrates the need for organizations implementing a secure messaging platform to truly teach patients how to use this tool and be explicit in communicating when and when not to use secure messaging. That said, the VA does inform its patients not to use secure messaging for urgent issues. Yet our findings demonstrate the difficulty inherent in educating patients on how to appropriately use new tools for communicating with their providers. Health care systems using secure messaging may need to revise their business rules to accommodate the need to respond to urgent messages. Many organizations' business rules are predicated on the assumption that no urgent or emergency messages will be sent via secure message. If even a small percentage of patients continue to use secure messaging for urgent issues, this assumption breaks down and introduces patient safety risks. One solution is to include a first message that must be viewed prior to sending a message that reads something like "If this is a medical emergency, dial 911." Increasing the number of staff available to handle secure messages quickly as opposed to 24 to 72 hours is another

potential solution. Accepting that patients may have a legitimate reason to use secure messaging for urgent matters (for instance, if the clinic phone lines are down and the patient is homebound) is another path forward. In this case, allowing the patient to flag their message as emergent or urgent (with definitions clearly labeled) could be a solution. Regardless, the problem of patients using secure messaging to communicate urgent issues remains and the potential patient safety risks can be serious.

Unresolved Issues, Patient Safety, and Patient Engagement

The finding that issues raised by patients went unresolved presents a challenge to both patient safety and patient engagement. Our analyses found that approximately 50% of the messages categorized as unresolved (6/11) retained that categorization following our medical record review. The potential implications for patient safety and patient engagement are clear. If a patient does not receive a response to a message, particularly if multiple messages receive no response, patient engagement could suffer. Patient engagement, or activation, is an important quality indicator for health care delivery organizations today. If a message contains an urgent issue and it is unresolved, then patient safety may be at risk. One of the remaining 6 unresolved messages in our data contained an urgent issue, which was surprising. Therefore, both patients and representatives of health care delivery systems need to be vigilant about using secure messaging to ensure patient messages are resolved and patient safety is not at risk.

Future Considerations

Secure messaging is one channel among many for communication between patients and members of their care team. However, little is known about what is unique about secure messaging as a patient-provider communication channel. Likewise, knowledge of potential harms introduced by features of secure messaging, such as asynchronous interaction and difficulty interpreting emotional cues via electronic communication channels, is limited. Future research studying the strengths and weaknesses of secure messaging should examine not only secure messaging as a new efficiency-enhancing communication channel for patients and their providers, but also the potential negative impacts on patient safety and patient engagement—particularly when patients' goals and intentions for secure messaging are misaligned with providers' goals and intentions for this communication platform.

Research questions that emerged from this analysis are as follows. How can health care organizations ensure secure messaging is not contributing to new, unanticipated patient safety concerns? Regarding urgent medical matters, is secure messaging a poor communication channel choice? If so, what are some effective strategies for communicating this to patients to avoid introducing new patient safety issues? If an issue a patient raises via secure message goes unresolved, are new patient safety issues introduced? Similarly, does a tone mismatch between patients and their provider teams result in decreased patient satisfaction or patient engagement? These questions tie back directly to our overarching theoretical frame of complexity science, which considers secure message use as a patient-provider interdependency that is interaction oriented and unpredictably dynamic, and to a forward-looking perspective linking improved secure message use with better patient satisfaction, engagement, and health outcomes.

Limitations

This study had several limitations that should be considered. Because of the qualitative nature of our study, we analyzed only 70 secure message threads. Given the volume of secure messages sent between patients and their providers today, this is a small number. However, our analysis included all secure messages sent in the time period in which the messages were sent, accounting for all secure messaging communication during that period of time. The purpose of this study was to identify and discuss new considerations for secure messaging as they may relate to key patient outcomes such as safety and engagement, as opposed to providing another detailed description of a large repository of secure messages. We believe the information management and uncertainty management categories and their potential to introduce unanticipated patient safety risks and engagement opportunities are novel contributions that add to the larger conversation of how to effectively use secure messaging platforms in health care delivery.

The research setting could be viewed as a limitation. The secure message triage model used in this VA is not uncommon; however, it may not always reflect the secure messaging implementation and use in other health care systems. While the VA is unique in many ways, the challenges it faces with regard to health IT adoption and use by patients and providers are similar to the challenges other health care delivery systems face.

These data were collected in 2013, and secure messaging communication practices may have evolved since then. We also acknowledge that the volume of messages has risen significantly, which may affect how patients and providers communicate with each other using this tool. Additionally, we acknowledge that the VA has worked since our data were collected on guidelines regarding the appropriateness of using secure messaging to address different issues.

We also acknowledge the limits of drawing inferences about patient safety and risks from viewing secure messages (and medical records) alone. This study did not measure the patient-provider relationship external to the messages. We also did not assess patient preferences about communication with their providers; thus, a tonal mismatch, for example, may not negatively affect the patient-provider relationship if both parties have an existing relationship that is strong and might anticipate or overlook tone mismatch.

Finally, this study focused on examining secure messages, and we acknowledge that potential safety concerns are not as compelling as actual patient safety lapses, or even observed near misses. Future research should take a step further in measuring patient safety risks in secure messaging and develop methods for identifying and verifying concerns raised by patients in secure messages and for acting on them in other processes of care experienced by the patient.

Conclusions

This analysis provides new insights into the complexity of secure messaging communication between patients and their outpatient care teams. Patients use secure messaging to exchange information, solve problems, build relationships, and make sense of their health or illness with their providers. Understanding the extent to which problems initiated via secure message go unresolved is an important piece of the puzzle for understanding the role of secure messaging in the patient-provider team communication toolbox and the potential for unpredictable negative impacts on secure messaging as a communication channel. Likewise, understanding the frequency with which patients are using secure messaging to communicate urgent medical matters is important, particularly given the potential risk to patient safety. Tone mismatches in care team response to patient secure message content is important to examine further because of their potential to negatively affect the patient-centered goals of health care organizations and the overall experience patients have with their health care providers. The patterns identified in this analysis shed light on potential patient safety concerns, particularly when using secure messaging to address some of the more complex issues patients are raising via secure messaging technologies. Finally, this study generated new questions for secure message use requiring additional examination.

Acknowledgments

This study was supported with resources and the use of facilities at the South Texas Veterans Health Care System, Veterans Evidence Based Research, Dissemination, and Implementation Center (VERDICT), San Antonio, Texas, USA. Support was also provided by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service. The views expressed in this paper are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US government.

Conflicts of Interest

None declared.

References

1. Cronin RM, Davis SE, Shenson JA, Chen Q, Rosenbloom ST, Jackson GP. Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. Appl Clin Inform 2015;6(2):288-304 [FREE Full text] [doi: 10.4338/ACI-2014-12-RA-0117] [Medline: 26171076]
2. Brooks RG, Menachemi N. Physicians' use of email with patients: factors influencing electronic communication and adherence to best practices. J Med Internet Res 2006;8(1):e2 [FREE Full text] [doi: 10.2196/jmir.8.1.e2] [Medline: 16585026]
3. Byrne JM, Elliott S, Firek A. Initial experience with patient-clinician secure messaging at a VA medical center. J Am Med Inform Assoc 2009;16(2):267-270 [FREE Full text] [doi: 10.1197/jamia.M2835] [Medline: 19074303]
4. Kittler AF, Carlson GL, Harris C, Lippincott M, Pizziferri L, Volk LA, et al. Primary care physician attitudes towards using a secure web-based portal designed to facilitate electronic communication with patients. Inform Prim Care 2004;12(3):129-138. [Medline: 15606985]
5. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
6. Miller DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
7. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
8. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]
9. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
10. Houston TK, Sands DZ, Jenckes MW, Ford DE. Experiences of patients who were early adopters of electronic communication with their physician: satisfaction, benefits, and concerns. Am J Manag Care 2004 Sep;10(9):601-608 [FREE Full text] [Medline: 15515992]
11. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]

12. Sittig DF. Results of a content analysis of electronic messages (email) sent between patients and their physicians. BMC Med Inform Decis Mak 2003 Oct 01;3:11 [FREE Full text] [doi: 10.1186/1472-6947-3-11] [Medline: 14519206]
13. McInnes DK, Shimada SL, Midboe AM, Nazi KM, Zhao S, Wu J, et al. Patient use of electronic prescription refill and secure messaging and its association with undetectable HIV viral load: a retrospective cohort study. J Med Internet Res 2017 Feb 15;19(2):e34 [FREE Full text] [doi: 10.2196/jmir.6932] [Medline: 28202428]
14. Nazi KM. Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. J Am Med Inform Assoc 2010;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]
15. Shimada SL, Hogan TP, Rao SR, Allison JJ, Quill AL, Feng H, et al. Patient-provider secure messaging in VA: variations in adoption and association with urgent care utilization. Med Care 2013 Mar;51(3 Suppl 1):S21-S28. [doi: 10.1097/MLR.0b013e3182780917] [Medline: 23407007]
16. Wolcott V, Agarwal R, Nelson DA. Is provider secure messaging associated with patient messaging behavior? Evidence From the US Army. J Med Internet Res 2017 Apr 06;19(4):e103 [FREE Full text] [doi: 10.2196/jmir.6804] [Medline: 28385681]
17. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente electronic health record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009 Apr;28(2):323-333 [FREE Full text] [doi: 10.1377/hlthaff.28.2.323] [Medline: 19275987]
18. Liederman EM, Morefield CS. Web messaging: a new tool for patient-physician communication. J Am Med Inform Assoc 2003;10(3):260-270 [FREE Full text] [doi: 10.1197/jamia.M1259] [Medline: 12626378]
19. Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care 2007 Jul;13(7):418-424 [FREE Full text] [Medline: 17620037]
20. Daft RL, Lengel RH. Organizational information requirements, media richness and structural design. Manage Sci 1986 May;32(5):554-571. [doi: 10.1287/mnsc.32.5.554]
21. Jarvenpaa SL, Knoll K, Leidner DE. Is anybody out there? Antecedents of trust in global virtual teams. J Manage Inf Syst 2015 Dec 08;14(4):29-64. [doi: 10.1080/07421222.1998.11518185]
22. Montoya-Weiss MM, Massey AP, Song M. Getting it together: temporal coordination and conflict management in global virtual teams. Acad Manage J 2001 Dec;44(6):1251-1262. [doi: 10.5465/3069399]
23. Straus SG. Getting a clue: the effects of communication media and information distribution on participation and performance in computer-mediated and face-to-face groups. Small Group Res 2016 Aug 18;27(1):115-142. [doi: 10.1177/1046496496271006]
24. Tang P, Black W, Young C. Proposed criteria for reimbursing eVisits: content analysis of secure patient messages in a personal health record system. : American Medical Informatics Association; 2006 Presented at: AMIA Annual Symposium; Nov 11-15, 2006; Washington, DC, USA p. 764.
25. White CB, Moyer CA, Stern DT, Katz SJ. A content analysis of e-mail communication between patients and their providers: patients get the message. J Am Med Inform Assoc 2004;11(4):260-267 [FREE Full text] [doi: 10.1197/jamia.M1445] [Medline: 15064295]
26. Bar Yam Y. Making Things Work: Solving Complex Problems in a Complex World. Cambridge, MA: Knowledge Press; 2004.
27. Cilliers P. Complexity and Postmodernism: Understanding Complex Systems. New York, NY: Routledge; 1998.
28. Kauffman S. At Home in the Universe: The Search for the Laws of Self-Organization and Complexity. Oxford, UK: Oxford University Press; 1995.
29. Zimmerman B, Lindberg C, Plsek P. Edgeware: Insights From Complexity Science for Health Care Leaders. Irving, TX: VHA Inc; 1998.
30. Lanham HJ, Leykum LK, McDaniel RR. Same organization, same electronic health records (EHRs) system, different use: exploring the linkage between practice member communication patterns and EHR use patterns in an ambulatory care setting. J Am Med Inform Assoc 2012;19(3):382-391 [FREE Full text] [doi: 10.1136/amiajnl-2011-000263] [Medline: 21846780]
31. Lanham HJ, Sittig DF, Leykum LK, Parchman ML, Pugh JA, McDaniel RR. Understanding differences in electronic health record (EHR) use: linking individual physicians' perceptions of uncertainty and EHR use patterns in ambulatory care. J Am Med Inform Assoc 2014;21(1):73-81 [FREE Full text] [doi: 10.1136/amiajnl-2012-001377] [Medline: 23698256]
32. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
33. McDaniel RR, Driebe DJ, Lanham HJ. Health care organizations as complex systems: new perspectives on design and management. Adv Health Care Manag 2013;15:3-26. [Medline: 24749211]
34. Lanham HJ, McDaniel RR, Crabtree BF, Miller WL, Stange KC, Tallia AF, et al. How improving practice relationships among clinicians and nonclinicians can improve quality in primary care. Jt Comm J Qual Patient Saf 2009;35(9):457-466. [Medline: 19769206]

35. Miller WL, Crabtree BF, McDaniel RR, Stange KC. Understanding change in primary care practice using complexity theory. J Fam Pract 1998;46:369-376. [Medline: 9597994]
36. Anderson R, Toles M, Corazzini K, McDaniel R, Colón-Emeric C. Local interaction strategies and capacity for better care in nursing homes: a multiple case study. BMC Health Serv Res 2014;14(1):1-17. [Medline: 24903706]
37. Albolino S, Cook R, O'Connor M. Sensemaking, safety, and cooperative work in the intensive care unit. Cogn Technol Work 2006 Dec 7;9(3):131-137. [doi: 10.1007/s10111-006-0057-5] [Medline: 26811110]
38. Zheng K, Ciemins E, Lanham H, Lindberg C. Examining the Relationship Between Health IT and Ambulatory Care Workflow Redesign (prepared by Billings Clinic under contract no. 290-2010-0019I-1). AHRQ Publication No. 12-0058-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2015:1-99.
39. Lanham HJ, Palmer RF, Leykum LK, McDaniel RR, Nutting PA, Stange KC, et al. Trust and reflection in primary care practice redesign. Health Serv Res 2015;51(4):1489-1514. [Medline: 26611650]
40. Leykum LK, Lanham HJ, Pugh JA, Parchman ML, Anderson RA, Crabtree BF, et al. Manifestations and implications of uncertainty for improving healthcare systems: an analysis of observational and interventional studies grounded in complexity science. Implement Sci 2014 Nov 19;9:165 [FREE Full text] [doi: 10.1186/s13012-014-0165-1] [Medline: 25407138]
41. McDaniel RR, Driebe D. editors. Uncertainty and Surprise in Complex Systems: Questions on Working With the Unexpected. New York, NY: Springer Science & Business Media; 2005.
42. U.S. Department of Veterans Affairs. Strategic Plan FY 2018-2024 strategic plan. Washington, DC: U.S. Department of Veterans Affairs; 2018 Feb 12. URL: https://www.va.gov/oei/docs/VA2018-2024strategicPlan.pdf [accessed 2018-06-04] [WebCite Cache ID 6zvVSdh6E]
43. Namey E, Guest G, Thairu L. Data reduction techniques for large qualitative data sets. In: Guest G, MacQueen K, editors. Handbook for Team-based Qualitative Research. Lanham, MD: AltaMira Press; 2008:137-161.
44. Ezzy D. Qualitative Analysis. Sydney, Australia: Routledge; 2013.
45. Strauss A. Qualitative Analysis for Social Scientists. Cambridge, UK: Cambridge University Press; 1987.
46. Shannon C. A mathematical theory of communication. Bell Syst Tech J 1948;27:623-656.
47. Simon HA. A behavioral model of rational choice. Q J Econ 1955;69(1):99-118.
48. Boisot M, Child J. Organizations as adaptive systems in complex environments: the case of China. Org Sci 1999 Jun;10(3):237-252. [doi: 10.1287/orsc.10.3.237]
49. Weick KE. The collapse of sense making in organizations: the Mann Gulch disaster. Adm Sci Q 1993 Dec;38(4):628. [doi: 10.2307/2393339]
50. Maguire S, Allen P, McKelvey B. Complexity management: introducing the SAGE handbook. In: Allen P, Maguire S, McKelvey B, editors. The SAGE Handbook of Complexity and Management. Los Angeles, CA: SAGE; 2011.
51. Dervin B. Information as non-sense; information as sense: the communication technology connection. In: Dervin B, Foreman-Wernet L, Launterbach E, editors. Sensemaking Methodology Reader: Selected Writings of Brenda Dervin. Cresskill, NJ: Hampton Press, Inc; 2003:293-308.
52. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk? J Am Med Inform Assoc 2013;20(6):1143-1149 [FREE Full text] [doi: 10.1136/amiajnl-2012-001208] [Medline: 23703826]
53. Johnson LW, Garrido T, Christensen K, Handley M. Successful practices in the use of secure e-mail. Perm J 2014 Jul;18(3):50-54 [FREE Full text] [doi: 10.7812/TPP/13-160] [Medline: 24937149]
54. Lewis MA, Dicker AP. Social media and oncology: the past, present, and future of electronic communication between physician and patient. Semin Oncol 2015 Oct;42(5):764-771 [FREE Full text] [doi: 10.1053/j.seminoncol.2015.07.005] [Medline: 26433557]
55. Ozkaynak M, Johnson S, Shimada S, Petrakis BA, Tulu B, Archambeault C, et al. Examining the multi-level fit between work and technology in a secure messaging implementation. AMIA Annu Symp Proc 2014;2014:954-962 [FREE Full text] [Medline: 25954403]
56. Popeski N, McKeen C, Khokhar B, Edwards A, Ghali WA, Sargious P, et al. Perceived barriers to and facilitators of patient-to-provider e-mail in the management of diabetes care. Can J Diabetes 2015 Dec;39(6):478-483. [doi: 10.1016/j.jcjd.2015.07.001] [Medline: 26409770]
57. Siva C, Lawlor K, Smarr K, Ge B, Fleming D. E-mail communication with patients: a survey of the American College of Physicians, Missouri Chapter. Mo Med 2011;108(2):136-140. [Medline: 21568238]
58. Ye J, Rust G, Fry-Johnson Y, Strothers H. E-mail in patient-provider communication: a systematic review. Patient Educ Couns 2010 Aug;80(2):266-273 [FREE Full text] [doi: 10.1016/j.pec.2009.09.038] [Medline: 19914022]

Abbreviations

IT: information technology
VA: US Department of Veterans Affairs

Edited by G Eysenbach; submitted 25.10.17; peer-reviewed by R Kitzmiller, J Lee; comments to author 30.11.17; revised version received 25.04.18; accepted 08.05.18; published 11.07.18.

Please cite as:
Lanham HJ, Leykum LK, Pugh JA
Examining the Complexity of Patient-Outpatient Care Team Secure Message Communication: Qualitative Analysis
J Med Internet Res 2018;20(7):e218
URL: http://www.jmir.org/2018/7/e218/
doi:10.2196/jmir.9269
PMID:29997107

Original Paper

Health Care Robotics: Qualitative Exploration of Key Challenges and Future Directions

Kathrin Cresswell[1], BSc, MSc, PhD; Sarah Cunningham-Burley[1], PhD; Aziz Sheikh[1], MSc, MD

Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Kathrin Cresswell, BSc, MSc, PhD
Usher Institute of Population Health Sciences and Informatics
University of Edinburgh
Level 2, Nine Edinburgh BioQuarter
9 Little France Road
Edinburgh,
United Kingdom
Phone: 44 7502 154328
Email: kathrin.beyer@ed.ac.uk

Abstract

Background: The emergence of robotics is transforming industries around the world. Robot technologies are evolving exponentially, particularly as they converge with other functionalities such as artificial intelligence to learn from their environment, from each other, and from humans.

Objective: The goal of the research was to understand the emerging role of robotics in health care and identify existing and likely future challenges to maximize the benefits associated with robotics and related convergent technologies.

Methods: We conducted qualitative semistructured one-to-one interviews exploring the role of robotic applications in health care contexts. Using purposive sampling, we identified a diverse range of stakeholders involved in conceiving, procuring, developing, and using robotics in a range of national and international health care settings. Interviews were digitally recorded, transcribed verbatim, and analyzed thematically, supported by NVivo 10 (QSR International) software. Theoretically, this work was informed by the sociotechnical perspective, where social and technical systems are understood as being interdependent.

Results: We conducted 21 interviews and these accounts suggested that there are significant opportunities for improving the safety, quality, and efficiency of health care through robotics, but our analysis identified 4 major barriers that need to be effectively negotiated to realize these: (1) no clear pull from professionals and patients, (2) appearance of robots and associated expectations and concerns, (3) disruption of the way work is organized and distributed, and (4) new ethical and legal challenges requiring flexible liability and ethical frameworks.

Conclusions: Sociotechnical challenges associated with the effective integration of robotic applications in health care settings are likely to be significant, particularly for patient-facing functions. These need to be identified and addressed for effective innovation and adoption.

(J Med Internet Res 2018;20(7):e10410) doi:10.2196/10410

KEYWORDS

robotics; health care; sociotechnical

Introduction

We are amid what has been described as the Fourth Industrial Revolution, where industries and sectors across the globe are being transformed using a variety of increasingly interconnected robotic applications [1]. These have demonstrably increased productivity, resource efficiency, and customer responsiveness in, for example, the manufacturing and retail sectors (see Figure 1) [2,3]. Amazon, for instance, now has a 100,000-robot fleet designed to navigate large warehouse spaces and pick items from shelves. This represents a 50% increase from the previous year such that robots now constitute around one-third of the workforce [4,5].

There is emerging policy interest in seeing a similar transition in health care; this is being fueled by the drive to improve the quality and safety of care while simultaneously controlling

expenditure [6]. Developments currently taking place have begun to replace individual aspects of human performance with robotic capabilities including precision (eg, surgical robots), logistic and mechanical tasks (eg, service robots), and complex cognitive tasks (eg, rehabilitation robots; see Figure 2 and Table 1) [7].

Deployments of robots in health care settings are likely to rise because of increasing technological capabilities, their reduced costs, and increasing pressure to curb costs. However, robots are potentially highly disruptive innovations, and it is therefore important to understand the sociotechnical challenges likely to be encountered as robots are deployed to find mitigating strategies [8-10]. Sociotechnical approaches to study the implementation of technology view social and technical factors as shaping each other over time. It is assumed that technologies are shaped by their social environments (eg, through designs being modified) but also that social environments are shaped by technological features (eg, when work practices of users change as a result of technology introduction).

Figure 1. Robotics in car manufacturing. Source: gyn9037/Shutterstock.com.

Figure 2. Robotics in health care. Source: Zapp2Photo/Shutterstock.com.

Table 1. Uses of health care robotics.

Type of device	Autonomous	Semiautonomous	Operational	Health care delivery, patient- and staff-facing
Service robots (eg, stock control, cleaning, delivery, sterilization)	✓	✓	✓	
Surgical robots		✓		✓
Telepresence robots (eg, screens on wheels)		✓		✓
Companion robots	✓			✓
Cognitive therapy robots	✓			✓
Robotic limbs and exoskeletons		✓		✓
Humanoids	✓		✓	✓

Such insights can support the development of an informed robotics strategy for health care that addresses these upcoming challenges (eg, by training staff and designing existing spaces appropriately), thus supporting the aim of transformation of health care through health information technology (HIT). To inform these important deliberations, we undertook an exploratory qualitative study to identify key sociotechnical challenges associated with introducing robotics in health care settings from the standpoint of key stakeholders.

Methods

Overview

We conducted an interview-based qualitative case study consulting stakeholders from various backgrounds and disciplines [11]. In doing so, health care robotics was conceptualized as the case. Other case studies currently in progress as part of a wider project exploring next generation technologies in health care settings include the integration of patient- and person-generated data with electronic health records (EHRs), innovative information infrastructures, and novel approaches to secondary data analysis.

Ethics and Permissions

This study received Institutional Review Board approval from the Centre of Population Health Sciences at the University of Edinburgh, United Kingdom. Participants gave written informed consent to participate, and transcripts were anonymized.

Sampling and Recruitment

Participants were sampled through Google searches using search terms relating to robotics and health care. We sampled purposefully for maximum variability ensuring presentation from a range of countries and professional backgrounds (including engineers, system developers, suppliers, academics, visionaries/futurists, users of robots in health care settings, and strategists) [12]. In line with the sociotechnical approach, the range of perspectives was expected to give important insights into the technical and social environments of robotic applications in health care settings. This sampling strategy was complemented by snowball sampling additional participants [13]. As our purpose was to develop a high-level overview, we did not specifically sample for individual users of applications in specific contexts.

Overall, we identified 68 participants. Of these, 42 were contacted through publicly available email addresses. The rest were sent invitations via LinkedIn through the account of the first author (KC). The initial email included an invitation to participate and an overview of the work. If participants expressed an interest (17/68 did), they were sent an information sheet and consent form. The remainder were sent a follow-up email approximately 2 weeks later (resulting in 9 additional responses). After initial discussions, 5 potential participants decided not to participate, mainly due to concerns surrounding signing the consent form and the interview being audio-recorded (although an option of not recording was offered). Industry representatives were not comfortable sharing potentially sensitive commercial information.

Data Collection

Interviews were conducted over Skype, digitally recorded, and transcribed verbatim by a professional transcriber. These ranged from 30 to 90 minutes, depending on the schedule of the participant and the number of issues they wanted to discuss. We explored the most promising areas surrounding health care robotics, their benefits and risks, anticipated and observed challenges, and potential ways to address these from a variety of technical and social angles. A sample interview guide can be viewed in Textbox 1.

We stopped recruiting participants when we reached thematic saturation (ie, when no new themes emerged during the concurrent analysis) [14]. To ensure that participant voices were reflected accurately, we performed member checking by sending the results to all participants and giving them the opportunity to comment on and correct any misunderstandings [15]. This resulted in minor clarifications to the results, consisting mainly of adding further details and context.

Data Analysis

Transcribed interviews were uploaded to NVivo 10 (QSR International Pty Ltd) software, which supports the management and interrogation of data and helps arrange qualitative data into meaningful headings and subheadings. We began the coding process as soon as interviews were transcribed to allow emerging findings to feed into future interviews; this involved sorting data into meaningful headings and subheadings for ongoing thematic analysis.

Textbox 1. Sample interview topic guide.

Vision surrounding robotics and automation in health care:

- Most promising developments to look out for, benefits
- What processes lend themselves best to automation?
- Any risks, issues that are particularly relevant to robotics
- Convergence of robotics, artificial intelligence, and big data analytics: how is the area of robotics defined?

Experiences of technological innovation in health care:

- Experiences and lessons learned
- User involvement in design
- Anything we can learn from other sectors?
- Which factors hinder developments, and how might these be addressed?

We approached the analysis with an initial coding framework based on the available empirical literature surrounding sociotechnical factors of technology implementation in health care settings [16]. The deductive components were as follows:

- Technological dimension (including technological features, technological infrastructures)
- Social/human dimension (including usability, human-technology interaction, attitudes)
- Organizational dimension (including organizational strategy, management, implementation)
- Macro-environmental dimension (regulation, legal, and ethical dimensions)

This allowed us to provide initial structure to our findings that remained close to the research question and sociotechnical perspective underpinning it. The coding framework was informed by our previous theoretical work surrounding the evaluation of sociotechnical systems [17]. In addition, we allowed new themes to emerge based on the frequency of occurrence and perceived significance (inductive component). During this process, we explored disconfirming evidence and carefully questioned our own (in some instances critical) assumptions about robotics.

In doing so, we carefully compared technological features, participant backgrounds, and insights into various sociotechnical aspects surrounding conceptualization, design, implementation, and adoption of technologies. Emerging themes were discussed and refined during regular meetings between the authors, paying particular attention to the intersection of technical and social factors in line with the sociotechnical lens.

Results

Overview

We interviewed a total of 21 participants (see Table 2). They came from a range of countries and academic, industry, and strategic backgrounds. Some, particularly academics, had mixed clinical backgrounds and had used or investigated robotic applications in health care contexts.

We identified a range of themes and subthemes, summarized in Textbox 2.

Overall, participants stated that the area of robotics in health care settings was still in its infancy and the move from paper-based to EHRs currently took strategic priority over investments in robotics. Specifically, the more novel developments surrounding humanoids were still seen to be a long way off in terms of routine deployment in health and care settings, while service robots were seen to hold the biggest short-term promise. However, it was also acknowledged that there was significant potential and the pace of developments as well as increasing convergence of applications meant that robotics was likely to become a routine aspect of health care delivery at some point.

> *I am quite taken by the fact of how quickly changes come about...in my lifetime as a surgeon in the late '80s we completely switched over a 2-year period from an open surgical approach to a minimal and key hole...* [Participant 2, surgeon, United States, male]

While some of the issues identified applied to all robotic uses outlined in Table 1, we also observed a hierarchy of features with increasing levels of sociotechnical complexity. For instance, semiautonomous operational applications tended to be viewed as presenting fewer sociotechnical challenges than autonomous care-facing functions. Further, there were subtle differences between participants from different backgrounds, with academics and strategists being slightly more critical, citing a wider range of challenges than commercial participants.

No Clear Pull From Professionals and Patients

There was a perception that concerns among the public, patients, and health care staff could hold back progress, leading to a lack of demand or acceptance for some robotic applications in health care settings. Attitudes were seen to be heavily influenced by negative publicity and modern science fiction.

> *...[patients] think when you say robot...you mean Terminator, so people are afraid...* [Participant 8, technologist, France, male]

Such negative attitudes were seen to be due to a range of factors. Some mentioned the clinician-patient relationship and patient trust as aspects of care that were perceived to require human

input. Therefore, applications seen to be performing tasks of a health care professional were viewed as particularly contested.

Table 2. Participant characteristics.

Participant number	Background	Country	Gender
1	Marketing: service robots	United States	Female
2	Surgeon: user of surgical robots	United States	Male
3	Academic: research into service robots	Norway	Male
4	Engineer: surgical robots	United States	Male
5	Futurist	United States	Female
6	Marketing: sterilization robots	Italy	Female
7	Academic, sociotechnical perspective	Switzerland	Male
8	Technologist: humanoids	France	Male
9	Academic: mainly surgical robots	United States	Male
10	Technologist	United Kingdom	Male
11	Engineer: telepresence robots	United States	Male
12	Academic	Sweden	Female
13	Strategist	Netherlands	Male
14	Journalist	United States	Male
15	Information technology consultant	United States	Male
16	Academic: informatics, rehabilitation, and surgical robots	United Kingdom	Male
17	Business development: humanoids	France	Male
18	Manager, robotics organization	France	Male
19	Academic: surgical robots	Australia	Male
20	Academic, ethicist	United Kingdom	Female
21	Academic, psychologist	United Kingdom	Female

Textbox 2. Themes identified in our work.

No clear pull from professionals and patients:

- Robots have negative publicity
- Lack of acceptance: trust is a social phenomenon and essential for health care
- Robots are transcending the human-machine interaction
- Lack of exposure to robots, particularly in Western cultures

Appearance of robots:

- Too robotic: psychological association with death, Terminator movie, fear of replacing human being
- Too human: expectations too high

Changes to the way health care work is organized and distributed:

- Changes to roles (replacing human capabilities versus augmenting them)
- Changes to workflows

New ethical and legal challenges:

- No existing liability and ethical frameworks
- Anticipating challenges will be crucial in the future
- Regulation is key to promote routine use

Purely service-based robots carrying out back-end functions were often seen as better suited to automation.

> *...to put your trust into a robot is still not there. I think a walker with robotic features is easier to adopt in the market by the people using it than a lifting robot.* [Participant 13, strategist, Netherlands, male]

From a health care provider point of view, negative attitudes were seen to be influenced by perceived threats to professional roles.

> *...a good anesthesiologist [costs] about $350 per hour, it's a heck of a wage, and the machine can be rented for about $150 so it's a lot more cost effective. That company have abandoned the product, not because it didn't function, it functioned extremely well. But it was very unpopular, and it had all sorts of doctor, patient unions and lobbying groups that had kittens about this idea of this robot that could basically put them all out of a job.* [Participant 5, futurist, United States, female]

Lack of exposure to robots was a major barrier to developing positive attitudes among patients and staff. This was seen to be since many existing applications such as pharmacy robots mainly operated in the back office, and there was a resulting fear of the unknown, particularly in Western cultures where robots are not routinely embedded in other aspects of everyday life.

> *...in Japan people believe that robots also have a kind of soul and that's why they approach robots as if they are like normal people. I believe in the rest of the world probably people...will be much more skeptical and I don't believe that people will accept that particularly caring for people will be performed by robots.* [Participant 7, academic, Switzerland, male]

Some participants suggested that public engagement campaigns, training of health care staff, and public dissemination of positive robotic case studies could help promote positive attitudes and acceptance of robotic applications among health care staff and patients.

Appearance of Robots

Humanoids presented a particularly interesting illustration of the tension between human hopes and expectations of robots and apprehension of their use in health care settings. They also represent an important sociotechnical example as human and technical dimensions blur in challenging and highly visible ways.

One reason identified in the interviews for humanoids not being very successfully integrated within health care settings was the contested nature of robotic appearances. On one hand, human features were seen as desirable in order to provide patients and staff with an experience of care as close to the real thing as possible.

> *We've tried to make it as approachable and friendly looking as possible because some people might think it's cold and now you're not having a direct person to person interaction in the flesh. So, we try to do our best to really make it as close as possible to the person being there, you know, with good audio, good video, the physical look of the robot.* [Participant 11, engineer, United States, male]

On the other hand, if robots were designed too human-like, there was significant apprehension of users reported, potentially being due to a fear of the robot replacing humans and imagined parallels with the Terminator vision of robots. This was particularly true for intimate tasks that often represented important aspects of the patient-provider relationship.

> *...there's a fear that the robot becomes almost like a near-human doppelganger that replaces the human being, because it has capabilities that we don't, so there's still this almost like mythical status of the robot that's certainly something that hovers around popular consciousness.* [Participant 21, manager, United Kingdom, female]

An additional undesired consequence was that when robots were designed as too human-like, they often fell short of human expectations of what they could do, resulting in disappointment and lack of engagement with and trust of the robot if it did not perform as expected.

> *...your expectations go up when you make robots human like.* [Participant 16, academic, United Kingdom, male]

Some also mentioned that the difficulty of placing humanoid robots firmly within either human or robotic categories was responsible for potential feelings of aversion. This was further exacerbated by a struggle to establish whether to perceive the robot as a friend or foe.

To address these problems, developers tended to design humanoid robots intentionally as non–human-like to ensure a visible demarcation between human and robotic features. This included, for example, designing them to roll on wheels rather than having legs or by designing them in the shape of an animal that most people had no experience interacting with (eg, a baby seal). This strategy was considered to be successful in promoting acceptability across contexts.

Disruption of the Way Work is Organized and Distributed

All participants acknowledged that the integration of robotic applications with existing health care professional work practices was important but difficult to achieve due to tensions between standardization through automation and the often-unpredictable nature of health care profession work.

The design of robotic applications that interacted with humans and spanned departmental and professional boundaries (ie, as autonomous robots) was seen as particularly problematic, as these transcended capabilities that were previously situated firmly within the human realm (eg, moving around, emotional support).

> *...when it comes into practice we all ran into problems. What if an elderly person is moving away from a robot, can it follow the elderly person? What if [the robot] falls, and it's a person with mild*

> *dementia. Is that person able to put the robot back on its feet again?* [Participant 13, strategist, Netherlands, male]

In contrast, robotic applications that were designated for particular uses (such as surgery, where they basically represented a sophisticated tool) or confined to back-office functions in controlled environments (such as pharmacy robots) were seen as less difficult to implement, as they had fewer challenging sociotechnical implications.

> *...this is why robots have been so successful in industry, like in car manufacturing, because they have these repetitive tasks and there are no humans in their way, they don't have to make decisions, they don't have to understand anything.* [Participant 14, journalist, United States, male]

It was therefore argued that to promote integration, robots should be viewed as augmenting human capabilities and empowering professionals in their role.

> *...when people talk about nurses and doctors and automation in a hospital, for example, the automation isn't about replacing the nurses and doctors, it's about augmenting their role so that they're more efficient so that they're not doing endless amounts of paperwork...they spend a bit more time with patients.* [Participant 10, technologist, United Kingdom, male]

This would, however, require some shift in skill sets toward supporting robotic capabilities and functions, particularly for lower skilled tasks. Envisioning and anticipating those changes was viewed as an important activity for educators, decision makers, and managers in health care settings.

New Ethical and Legal Challenges

Robotic applications engender new ethical and legal challenges surrounding their use in highly human social settings, and interviewees gave many examples. Some of these tackled the physical environment.

> *There was one lady who got trapped in an elevator together with one of the robots, and another one got run over.* [Participant 3, academic, Norway, male]

Others described psychological challenges such as the perceived risk of becoming too emotionally attached to a robot (particularly in care settings where patients are vulnerable).

> *...if you look at the target audience this will be vulnerable people, disabled people, sick people, the elderly...so it's important that we have robots that do not transport a feeling that is not real, like companionship robots, for example. They should be designed in a way that it's always clear that it's always a robot and not a substitute for a human.* [Participant 20, academic, United Kingdom, female]

Ethical dimensions surrounding nonuse of technology were also mentioned. These included issues of whether health care professionals should be forced to use a robotic application if this were a safer alternative than human-delivered care.

> *I'm particularly trying to answer the questions like if we show that you can do something more safely with the robot does that mean that people should use the robots if they know there's a safer alternative...should they be forced to use a robot assistant because they know it's a safer way of doing it...* [Participant 19, academic, Australia, male]

Some had begun developing ethical frameworks for robotic applications. A defining characteristic was that both human and machine perspectives were represented so that the guiding principles were both machine logic and human logic (including their reaction to machine behavior), implying that a new sociotechnical approach to HIT ethics is developing.

> *...the idea is that the framework is understandable by both humans and machines so that if a machine needs guidance, a human can work through the framework and figure out where it got stuck and make a judgment call or vice versa. Machines can begin to understand how humans themselves are making a certain decision and provide guidance or insight into that.* [Participant 5, futurist, United States, female]

Additionally, interviewees noted that there was a lack of clear, established liability rules surrounding robotics, made all the more problematic given the perceived hype surrounding robotics and a certain keenness of getting these into use quickly. This meant that when accidents happened (such as robots running over humans), these often had to be solved ad hoc, further contributing to negative public attitudes and inhibiting innovation.

Participants suggested that a more deliberative approach was needed to create clear liability rules surrounding product and consumer safety across different settings in which robotic applications were used, including health care. However, it was seen to be crucial to find a balance between developing overarching rules and allowing innovation to flourish.

Discussion

Principal Findings

Although there has been substantial technological progress in the field of health care robotics, robot integration into health care settings is likely to be far from straightforward. We have identified several concerns that are often shaped by preconceptions surrounding the appearance of robotic applications and associated (often conflicting) desires for human and technological features. In addition to these negative attitudes that result in a lack of user pull and demand, robotics also does and will change the way health care work is organized and distributed with some applications augmenting and others replacing human labor. These changes require new ethical and liability frameworks as new situations may emerge that blur the line between human responsibility and technological autonomy.

Comparison With Prior Work

In undertaking this study, we have elicited the perspectives and experiences of stakeholders from various international settings to bring together knowledge and deliberate on potential future challenges of implementing and optimizing robotic applications

in health care settings. We have identified sociotechnical challenges associated with various technological features. This builds on previous work focused on specific systems already being used in specific settings [18-21]. Our focus, in line with our uses (Table 1), was on different aspects of robotic hardware function. Although these were necessarily combined with some software capability including artificial intelligence, software was not the focus of our work.

There is an increasing recognition that sociotechnical considerations are important when considering technological applications including robotics [22,23] but only a limited number of studies have examined such issues with regard to robotic applications in health care [24-26]. Where it exists, primary research has concentrated on technologies in specific environments, including some in health care [27,28]. However, when compared to other HIT, autonomous applications (such as humanoids) present specific sociotechnical challenges because social and technical dimensions are progressively, visibly, and disruptively interconnected. As a result, there is a danger that these sociotechnical challenges will lead to an increasing range of problems integrating robotic applications within particularly human-dense social environments such as health care.

Ethical dimensions surrounding robotics, especially relating to trust and acceptance, have received relatively high levels of attention, perhaps due to perceived negative public attitudes surrounding robotic systems [29-31]. Our work has supported existing research highlighting that these issues pose important sociotechnical barriers to progress. Humans must renegotiate their roles within increasingly technological environments, and this negotiation is characterized by a conceptual struggle between a desire for progress and an apprehension toward the increasingly human side of machines.

Although important as a subject of ongoing debate, these issues are unlikely to ever be fully resolved. Some have found that trust and positive attitudes toward robotic applications can be promoted through exposure [28,32], and exposure is likely to be key in going forward. As robotic applications become more visible in everyday environments, they are likely to become more acceptable in health care settings. Lack of exposure is likely to be a transient issue as there are now many examples in other industries and countries where robots and humans routinely work alongside each other.

Limitations

The response rate to interview invitations was low (only 21 out of 68 individuals agreed to be interviewed), in part reflecting concerns about disclosing commercially sensitive information. We may therefore have missed some important considerations (despite having achieved thematic saturation within our sample), particularly from cultures that have integrated robotics in everyday life (eg, Japan). Additional factors that are likely to have shaped the sampling of respondents include the presence on Google and LinkedIn, access to Skype, English language facilities, and the Google search methods employed by the researcher. We therefore necessarily explored the views of those who were visible and vocal in relation to health care robotics in English media. Although this was appropriate for gaining a high-level overview into an underexplored topic, it also means that our results are likely to have missed the perspectives of certain user groups (eg, health care professionals and patients with or without the experience of robotics). This may have led to a lack of insight into the acceptability of specific applications. Such work is important going forward as many of the challenges identified are heavily dependent on individual settings, technologies, and contexts. Moreover, we acknowledge that we have only skimmed the surface of exploring ethical, legal, and policy dimensions of robotic applications in health care settings, and this would certainly be a fruitful area for further in-depth research. There was also a clear gender imbalance toward male respondents in our sample, perhaps due to the fact that experts in this area are predominantly male.

Implications for Research, Policy, and Practice

We have begun charting the range of sociotechnical challenges that are likely to test the routine integration and optimization of robotics into health care settings. We summarize these along with possible ways to address them in Table 3.

Although there is a large literature base addressing the promises of robotics, this is limited to applications other than health care or specific health care applications such as surgery [1,33]. There is a need for empirical investigations into potential challenges and unintended consequences of such technologies in health care settings.

New ethical and regulatory frameworks are now needed that are nimble enough to keep up with changing environments and the increase in and convergence of robotic functionality. This may need to involve training a new generation of professionals who specialize in high-risk settings such as health care because existing regulations simply cannot keep up with the pace of technological advancements. Work may also need to involve drawing on ongoing efforts in other industries where these challenges have begun to be addressed. Health care robotics is an emerging field that will need inclusive, designated working groups at national and international levels because many functions are patient- and staff-facing and humans and machines need to coexist and collaborate in high-risk environments.

Table 3. Sociotechnical challenges identified with suggested strategies.

Identified challenge	Suggested strategy
No clear pull from professionals and patients	Establish an accessible empirical evidence base associated with specific functionalities; communication of benefits and challenges
Appearance of robots	Closer working relationships between developers, psychologists, users, and human-centered design specialists
Changes to the way health care work is organized and distributed	Prospective longitudinal evaluation of the implementation, adoption, and optimization of technologies
New ethical and legal challenges	Development of new ethical and regulatory frameworks that are flexible enough to keep up with changing environments and robotic functionality

Robotics in designated controlled environments (such as service robots) are likely to be less problematic and bring the highest gains in the short term because they present a limited number of sociotechnical challenges compared with applications that blur social and technical dimensions (eg, humanoids).

Conclusions

Sociotechnical challenges surrounding the implementation of robotics in health care settings are significant, although these are likely to vary with different robotic applications and in different cultural contexts. These challenges need to be anticipated and, if possible, proactively addressed. Health care settings are characterized by their care work; the provocation is to preserve and intensify or augment this within an increasingly automated and technological environment. This can only be done if we anticipate challenges associated with new technologies and systematically address them as we integrate them within existing social orders. Our research should be seen as a stepping stone to stimulate wider discussions surrounding these challenges. It can also help to guide health care organizations and policy makers as they make important strategic decisions associated with purchasing, developing, and deploying robotic applications.

Acknowledgments

KC is supported by a Chief Scientist Research Grant, and AS is supported by the Farr Institute. We gratefully acknowledge the input of Rosemary Porteous, who transcribed the interviews, and all participants for giving their time.

Conflicts of Interest

None declared.

References

1. Siciliano B, Khatib O, editors. Springer Handbook of Robotics. New York: Springer; 2016.
2. UK-RAS Network. Manufacturing robotics: the next robotic industrial revolution URL: http://hamlyn.doc.ic.ac.uk/uk-ras/sites/default/files/UK_RAS_wp_manufacturing_web.pdf [accessed 2018-06-18] [WebCite Cache ID 70Hag6iBZ]
3. The King's Fund. Carter review on productivity in NHS hospitals URL: https://www.kingsfund.org.uk/projects/carter-review-productivity-nhs [accessed 2018-06-19] [WebCite Cache ID 70Har0jby]
4. Business Insider. 2017 Jan 03. Amazon now has 45,000 robots in its warehouses URL: http://uk.businessinsider.com/amazons-robot-army-has-grown-by-50-2017-1 [accessed 2018-06-18] [WebCite Cache ID 70Hb2n8XW]
5. The New York Times. 2017 Sep 10. As Amazon pushes forward with robots, workers find new roles URL: https://www.nytimes.com/2017/09/10/technology/amazon-robots-workers.html [accessed 2018-06-18] [WebCite Cache ID 70HbFJSYu]
6. Bates DW, Gawande AA. Improving safety with information technology. N Engl J Med 2003 Jun 19;348(25):2526-2534. [doi: 10.1056/NEJMsa020847] [Medline: 12815139]
7. Van der Loos H, Reinkensmeyer D, Guglielmelli E. Rehabilitation and health care robotics. In: Siciliano B, Khatib O, editors. Springer Handbook of Robotics. New York: Springer; 2016:1685-1728.
8. Coiera E, Ash J, Berg M, Principal A. The unintended consequences of health information technology revisited. IMIA Yearbook 2016 Nov 10:163-169. [doi: 10.15265/IY-2016-014]
9. Bijker W. Toward a Theory of Sociotechnical Change: Of Bicycles, Bakelites, and Bulbs. Cambridge: MIT Press; 1997.
10. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999 Aug;55(2):87-101. [Medline: 10530825]
11. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol 2011 Jun 27;11:100 [FREE Full text] [doi: 10.1186/1471-2288-11-100] [Medline: 21707982]
12. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. 4th Edition. Thousand Oaks: Sage Publications; 2015.
13. Biernacki P, Waldorf D. Snowball sampling: problems and techniques of chain referral sampling. Sociological Methods Res 1981;10(2):141-163. [doi: 10.1177/004912418101000205]

14. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]
15. Creswell J, Miller D. Determining validity in qualitative inquiry. Theory Practice 2000;39(3):124-130 [FREE Full text]
16. 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]
17. Cresswell K, Sheikh A. The NHS Care Record Service (NHS CRS): recommendations from the literature on successful implementation and adoption. Inform Prim Care 2009;17(3):153-160 [FREE Full text] [Medline: 20074427]
18. Leifer L, Toye G, Van der Loos M. Tele-service-robot: integrating the socio-technical framework of human service through the InterNet-world-wide-web. Robotics Autonomous Syst 1996;18(1-2):117-126. [doi: 10.1016/0921-8890(96)00019-X]
19. Garmann-Johnsen N, Mettler T, Sprenger M. Service robotics in healthcare: a perspective for information systems researchers? 2014 Presented at: Thirty Fifth International Conference on Information Systems; 2014; Aukland.
20. Jones K, Cherry B, Harris D, Sridharan M. Formative analysis of aging in place: implications for the design of caregiver robots. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2017 Sep 28;61(1):1145-1145. [doi: 10.1177/1541931213601770]
21. Cao G, Rogers G. Robotics in health care: HF issues in surgery. In: Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. Mahwah: Lawrence Earlbaum & Associates; 2007:411-421.
22. Midler C, Charue F. A French-style sociotechnical learning process: the robotization of automobile body shops. In: Kogut B, editor. Country Competitiveness. Technology and the Organizing of Work. Oxford: Oxford University Press; 1993:5-75.
23. Susman G, Chase R. A sociotechnical analysis of the integrated factory. J Applied Behav Sci 1986;22(3):257-270. [doi: 10.1177/002188638602200306]
24. Van Oost E, Reed D. Towards a sociological understanding of robots as companions. 2010 Presented at: Human Robot Personal Relationships: Third International Conference; 2010; Leiden p. 11-18.
25. Oborn E, Barrett M, Darzi A. Robots and service innovation in health care. J Health Serv Res Policy 2011 Jan;16(1):46-50. [doi: 10.1258/jhsrp.2010.010001] [Medline: 20729249]
26. Barrett M, Oborn E, Orlikowski W, Yates J. Boundary Relations: Technological Objects and the Restructuring of Workplace Boundaries. Cambridge: University of Cambridge; 2007.
27. Nyssen A, Blavier A. Investigating expertise, flexibility and resilience in socio-technical environments: a case study in robotic surgery. Resilient Health Care 2013:97-111 [FREE Full text]
28. Alaiad A, Zhou L. The determinants of home healthcare robots adoption: an empirical investigation. Int J Med Informatics 2014;83(11):825-840. [doi: 10.1016/j.ijmedinf.2014.07.003]
29. Lichocki P, Billard A, Kahn P. The ethical landscape of robotics. IEEE Robotics Automation Magazine 2011;18(1):39-50 [FREE Full text]
30. Crnkovic G, Çürüklü B. Robots: ethical by design. Ethics Inform Technol 2012;14(1):61-71. [doi: 10.1007/s10676-011-9278-2]
31. Broadbent E, Stafford R, MacDonald B. Acceptance of healthcare robots for the older population: review and future directions. Int J Soc Robotics 2009;1(4):319-330. [doi: 10.1007/s12369-009-0030-6]
32. Attitudes towards the impact of digitisation and automation on daily life.: European Commission; 2017 Apr 10. URL: https://ec.europa.eu/digital-single-market/en/news/attitudes-towards-impact-digitisation-and-automation-daily-life [accessed 2018-06-19] [WebCite Cache ID 70Ha2otJ7]
33. Van Wynsberghe A. Healthcare Robots: Ethics, Design and Implementation. London: Routledge; 2016.

Abbreviations

HIT: health information technology
EHR: electronic health record

Edited by G Eysenbach; submitted 14.03.18; peer-reviewed by B Kaplan, JM Cogollor, B Arnoldussen, A Cyr; comments to author 02.05.18; revised version received 04.05.18; accepted 05.05.18; published 04.07.18.

Please cite as:
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
URL: http://www.jmir.org/2018/7/e10410/
doi:10.2196/10410
PMID:29973336

Original Paper

Web-Based Intervention Using Behavioral Activation and Physical Activity for Adults With Depression (The eMotion Study): Pilot Randomized Controlled Trial

Jeffrey D Lambert[1], PhD; Colin J Greaves[2], PhD; Paul Farrand[3], PhD; Lisa Price[4], PhD; Anne M Haase[5], PhD; Adrian H Taylor[6], PhD

[1]Institute for Health Research, Primary Care, University of Exeter Medical School, Exeter, United Kingdom
[2]School for Sport, Exercise and Rehabilitation, University of Birmingham, Birmingham, United Kingdom
[3]Clinical Education, Development and Research, Psychology, University of Exeter, Exeter, United Kingdom
[4]Sport and Health Sciences, University of Exeter, Exeter, United Kingdom
[5]Centre for Exercise, Nutrition and Health Sciences, University of Bristol, Bristol, United Kingdom
[6]Faculty of Medicine & Dentistry, University of Plymouth, Plymouth, United Kingdom

Corresponding Author:
Jeffrey D Lambert, PhD
Institute for Health Research
Primary Care
University of Exeter Medical School
Smeall Building
Heavitree Roadd, University of Exeter
Exeter, EX1 2LU
United Kingdom
Phone: 44 1392725705
Email: j.d.lambert@exeter.ac.uk

Abstract

Background: Physical activity is a potentially effective treatment for depression and depressive relapse. However, promoting physical activity in people with depression is challenging. Interventions informed by theory and evidence are therefore needed to support people with depression to become more physically active. eMotion is a Web-based intervention combining behavioral activation and physical activity promotion for people in the community with symptoms of depression.

Objective: The objectives were to assess the feasibility and acceptability of delivering eMotion to people in the community with symptoms of depression and to explore outcomes.

Methods: Participants with elevated depressive symptoms were recruited from the community through various methods (eg, social media) and randomized to eMotion or a waiting list control group for 8 weeks. eMotion is an administratively supported weekly modular program that helps people use key behavior change techniques (eg, graded tasks, action planning, and self-monitoring) to re-engage in routine, pleasurable, and necessary activities, with a focus on physical activities. Feasibility data were collected that included the following: recruitment and trial retention rates; fidelity of intervention delivery, receipt, and enactment; and acceptability of the intervention and data collection procedures. Data were collected for the primary (depression) and secondary outcomes (eg, anxiety, physical activity, fidelity, and client satisfaction) at baseline and 2 months postrandomization using self-reported Web-based questionnaires and accelerometers. Delivery fidelity (logins, modules accessed, time spent) was tracked using Web usage statistics. Exploratory analyses were conducted on the primary and secondary outcomes.

Results: Of the 183 people who contacted the research team, 62 were recruited and randomized. The mean baseline score was 14.6 (SD 3.2) on the 8-item Patient Health Questionnaire depression scale (PHQ-8). Of those randomized, 52 participants provided accelerometer-recorded physical activity data at baseline that showed a median of 35.8 (interquartile range [IQR] 0.0-98.6) minutes of moderate-to-vigorous physical activity (MVPA) recorded in at least 10-minute bouts per week, with only 13% (7/52) people achieving guideline levels (150 minutes of MVPA per week). In total, 81% (50/62) of participants provided follow-up data for the primary outcome (PHQ-8), but only 39% (24/62) provided follow-up accelerometer data. Within the intervention group, the median number of logins, modules accessed, and total minutes spent on eMotion was 3 (IQR 2.0-8.0), 3 (IQR 2.0-5.0), and 41.3 (IQR 18.9-90.4), respectively. Acceptability was mixed. Exploratory data analysis showed that PHQ-8 levels were lower

for the intervention group than for the control group at 2 months postrandomization (adjusted mean difference −3.6, 95% CI −6.1 to −1.1).

Conclusions: It was feasible to deliver eMotion in UK communities to inactive populations. eMotion has the potential to be effective and is ready for testing in a full-scale trial. Further work is needed to improve engagement with both the intervention and data collection procedures.

Trial Registration: ClinicalTrials.gov NCT03084055; https://clinicaltrials.gov/ct2/show/NCT03084055 (Archived by WebCite at http://www.webcitation.org/6zoyM8UXa)

(J Med Internet Res 2018;20(7):e10112) doi:10.2196/10112

KEYWORDS

psychological therapy; mood; anxiety; exercise; eHealth; feasibility; acceptability

Introduction

Depression has a significant detrimental impact on individuals and their families as well as being associated with increased utilization of health services and reduced productivity at work. In the UK, the cost to the economy due to sickness absence, staff turnover, benefits, fall in tax revenue, and costs to the National Health Service (NHS) is estimated to be between £74 billion and £99 billion per year [1]. Depression is also associated with a range of major physical illnesses (which are also associated with physical inactivity), including diabetes [2,3], coronary heart disease [4], and obesity [5].

Physical activity (PA) has been shown to be effective in treating [6-8] and preventing [9] depression and is often cited by patients as their preferred treatment option [10,11]. A recent systematic review of randomized controlled trials (RCTs) found PA to be more effective than control conditions for reducing depression (standardized mean difference [SMD] −0.62) and just as effective as pharmacotherapy (SMD −0.11) and psychological therapies (SMD −0.03) [6]. PA also has the potential to reduce depressive relapses [12], improve anxiety symptoms [13], and help prevent many physical health problems associated with depression such as cardiovascular disease, cancer, and diabetes [14]. Despite these benefits, PA is perceived as difficult to prescribe compared with medication [15]. Further studies are therefore needed to understand how and whether PA can be cost-effectively used to improve outcomes for people with depression, especially in those who are relatively inactive [16].

Behavioral activation (BA) is an evidenced-based psychological therapy. BA focuses on reducing an individual's exposure to sources of negative reinforcement (ie, short-term relief from avoiding burdensome activities) while increasing experiences of positive reinforcement (ie, social and personal activities that bring pleasure and achievement), leading to reduced avoidant type-behaviors in the future [17-19]. A meta-analysis of 26 RCTs (n=1524) conducted on adults with depression found BA to be superior to usual care, wait list, or placebo control conditions (SMD −0.74) and medication (SMD −0.42) for reducing depression [20]. The treatment rationale of BA shares behavior change techniques (BCTs) [21] with interventions promoting PA (eg, goal setting) [22]. By subtly shifting the behavioral emphasis, the treatment rationale of BA could therefore provide a useful delivery mechanism for promoting PA in people with depression, capitalizing on the dual benefits of PA and BA. Furthermore, sustained PA may reduce the relapse rate associated with BA alone due to its inherent mood enhancing and long-term benefits [12].

Recent studies have examined the feasibility of delivering a combined BA and PA intervention (BAcPAc) within the context of existing mental health services [22,23]. However, BAcPAc was difficult to implement in overstretched services, it was hard to recruit patients, and problems of fidelity were observed, with providers not delivering the treatment as intended [23]. Web-based interventions could provide a useful way of overcoming these limitations by delivering such interventions outside of existing services, recruiting directly from the community, and standardizing fidelity [24,25]. Furthermore, Web-based interventions could have additional benefits by being scalable, cheap, and accessible [26]. Importantly, Web-based interventions can also be used to provide support to people experiencing depression in UK communities who do not seek help from health services due to social stigma and identity conflict [27]. However, despite the potential benefits associated with Web-based interventions, there remains a paucity of studies that have attempted to deliver Web-based interventions promoting PA for depression [28-31]. Furthermore, none of these studies have explored the feasibly of delivering a Web-based intervention combining BA and PA for people with depression.

The purpose of the present study was to examine the feasibility of delivering a Web-based intervention (eMotion), combining PA with BA, for people with depression and to explore its effects on depression and PA. The key aims of the study were to: explore participant recruitment and attrition rates throughout the study; explore the feasibility and acceptability of data collection and study procedures; examine baseline characteristics of the recruited sample, including levels of PA; explore the fidelity of delivery, receipt, and enactment (use of techniques) of eMotion and its acceptability to participants; and estimate and explore the variance in key outcomes (depression scores and PA).

Methods

Trial Design

Description of Trial Design

The eMotion trial was a 2-arm, individually randomized, parallel-group pilot RCT with a nested process evaluation

(ClinicalTrials.gov Identifier: NCT03084055). The study was reported in accordance with Consolidated Standards of Reporting Trials (CONSORT-EHEALTH) recommendations for reporting of RCTs of eHealth interventions [32] and the Template for Intervention Description and Replication (TIDieR) recommendations on reporting of behavior change interventions [33].

Important Changes to Methods After Trial Commencement

The eMotion trial provided "minimal contact" administrative support [34] at week 2 of the intervention to provide the participant with a rationale for the use of self-help materials and check-ins related to progress, but with no focus on any clinical or behavior change issues. This support was initially intended to be provided by an independent "supporter." However, due to resource issues, this support was provided by the lead author.

Participants

Eligibility Criteria for Participants

Participants were eligible for the study if they were more than 18 years old, were living in UK, had at least moderate depressive symptoms [defined as scoring at least 10 on the 8-item Patient Health Questionnaire depression scale (PHQ-8)], had access to the internet and were computer literate, reported being able to walk continuously and unaided for a minimum of 5 min, and provided informed consent to participate. Eligible adults were recruited from the community via advertisements in weekly newspapers, social media (eg, Facebook support groups, Twitter), and through banners on websites relating to mental health problems. All adverts contained the primary investigator's contact details. Potential participants did not need to be referred by a general practitioner or mental health care practitioner. After contacting the lead author by phone or email, potential participants were sent the participant information sheet (PIS), consent form, and a link to the Web-based screening questionnaire via email. At this point, they were informed that they could withdraw from the study at any time without consequence or being obliged to provide a reason. Once participants read the PIS, they were asked to complete the Web-based consent form (indicating consent using a checkbox) followed by a screening questionnaire used by the lead author to assess participant eligibility. After screening, the lead author contacted the participant via phone to clarify the study procedures and provide instructions for wearing the wrist-worn accelerometers sent by post. The participant was instructed to wear the accelerometer for 7 days and return it in a pre-stamped addressed envelope. However, participant refusal to wear an accelerometer did not preclude randomization. Further baseline measures were then administered via a separate Web-based questionnaire. Participants were not paid for their involvement in the study.

The eMotion Intervention

The eMotion intervention is a Web-based course that provides people with access to an evidence-based treatment based on BA with added PA promotion. People with depression commonly reduce activities that they perceive as burdensome, making less effort to do things they may have previously enjoyed. By avoiding such activities, people with depression experience temporary relief that then negatively reinforces the likelihood of avoiding further activities. However, avoiding such activities has a long-term cost because it reduces the opportunities for positive reinforcement that occurs when people engage in social and personal activities that bring them pleasure and achievement [17-19]. PA is often avoided by people with depression, but it has the potential to provide additional anti-depressive benefits as well as added health benefits.

Through a series of steps delivered in a week-by-week modular fashion (Table 1), eMotion teaches people how to re-establish daily routines, increasing activities that provide positive reinforcement while reducing negative reinforcement. eMotion comprises 13 modules (1 introduction module, 8 weekly modules, 1 generic problem-solving module, and 3 unlockable modules) consisting of visual content with an audio voiceover triggered when each slide opens. Printable, interactive worksheets, and emails were also included, with links to the slides to allow downloading to a personal computer or another device (eg, tablet or smartphone). Automated reminder emails were also sent once a week by the eMotion program following registration. Where possible, brief administrative/motivational support via a 10-minute phone call was provided at week 2.

Key content related to the rationale of BA was front loaded in the introduction, week 1, and week 2 and was considered the "minimum dose." The remaining weekly modules (weeks 3-8) were shorter and designed to support people to review and update their plans. The intervention supports "effective engagement" and self-regulation by encouraging people to review their plans weekly, irrespective of whether or not they continue to login (eg, using their own diaries).

eMotion Development

The content for eMotion was developed by the study authors (JDL, CJG, PF, AMH, and AHT). The Living Life to the Full Web-based platform was used to host the intervention. eMotion was adapted from the BAcPAc intervention [22,23] using the Centre for eHealth Research and Disease Management (CeHReS) roadmap [35]. The CeHReS roadmap is intended to help the planning, coordination, and execution of the participatory development process of eHealth. In eMotion, this involved using patient and public involvement, usability testing, and a structured literature search. A full description of the eMotion intervention and its developmental process was previously provided [36].

Control (Waiting List)

Participants in the control group did not receive the eMotion intervention but were able to access usual care as normal. After data collection at the 2-month time point, participants were given access to and instructions for using eMotion. No eMotion facilitator support was provided, although participants could contact the lead researcher for (nonclinical) support if they experienced any difficulties using the intervention.

Table 1. eMotion structure.

Steps	Content	Module in eMotion
Step 1: Provide a rationale	People are provided with a full and comprehensive rationale for behavioral activation and PA[a], including reference to the interaction of physiological, behavioral, and cognitive emotional symptoms, the role of avoidance in maintaining low mood and the idea of routine, pleasurable, necessary, and activities (including PA).	Introduction and Week 1
Step 2: Identify activities	People are helped to identify routine, pleasurable, and necessary activities (including PA)—things that they would like to do but have usually stopped doing since they became depressed.	Weeks 2-8
Step 3: Make a hierarchy of activities	People are helped to organize the activities into a hierarchy of difficulty—most difficult, medium difficult, and easiest. People should include some of each type of routine, pleasurable, and necessary activity (including PA).	Weeks 2-8
Step 4: Plan some activities	eMotion helps people to schedule some avoided activities into their week, to specify a mixture of routine, pleasurable, and necessary activities (including PA). These should be initially identified from the "easiest" category of their hierarchy from step 3). Activities should be detailed precisely: what, where, when, and with whom.	Weeks 2-8
Step 5: Implement activities	People are encouraged to undertake the planned activities. The principle of grading activities and using a mixture of routine, pleasurable, and necessary activities (including PA) should be followed. People should record if they accomplished the planned activity.	Weeks 2-8
Step 6: Review progress	People are encouraged to reflect on their progress, congratulating themselves for success and overcoming any problem-solving difficulties experienced during implementation. People may make sporadic progress and activities may not go as planned.	Weeks 3-8

[a]PA: Physical activity.

Feasibility Outcomes

Recruitment and Attrition

Participant recruitment rates were derived by calculating the absolute number of people randomized in the trial relative to those who expressed an initial interest in the study. Participant attrition was defined as the percentage of randomized participants who began the intervention but failed to provide primary outcome data (PHQ-8) at the 2-month data collection point.

Feasibility of Data Collection

Feasibility of data collection was explored by assessing the percentage of Web-based screening, baseline, and 2-month postrandomization questionnaires that were completed, as intended. For accelerometers, the percentage of devices that were returned at baseline and at 2 months follow-up as well as the amount of valid wear time were assessed. Reasons for any missing data were tabulated, where available.

Primary Outcome for the Planned Future Trial

Depression

The PHQ-8 was delivered at screening and at 2 months postrandomization using a Web-based self-completed version of the questionnaire. The PHQ-8 is a freely available 8-item self-report measure based on the symptoms of depression described in the Diagnostic and Statistical Manual for Mental Disorders (DSM-IV). It measures the frequency of depressive symptoms over the preceding 2-week period. A score of at least 10 on the PHQ-8 has a positive likelihood ratio of 28 for detecting major depression (ie, a patient with any depressive disorder is 28 times more likely to have a PHQ-8 score of 10-24 than someone without a depressive disorder) [37]. Each item is rated on a scale of 0-3, producing a range of scores from 0-24 (0-4=no depression, 5-9=mild depression, 10-14=moderate depression, 15-19=moderately severe depression, and 20-24=severe depression). The PHQ-8 has good validity, reliability, sensitivity, and specificity [37] and has been used in previous Web-based intervention studies of low mood and depression [38].

Secondary Outcomes for the Planned Future Trial

Objective Physical Activity

GENEActiv accelerometers (Activeinsights Ltd., Kimboloton, Cambs, UK) were used to record PA at baseline and at 2-months postrandomization. The GENEActiv is a small wrist-worn device that measures and records acceleration. It was set to record at 100Hz. Data were downloaded using the GENEActiv PC software (version 2.9) and processed in R (https://cran.r-project.org/) using package GGIR (version 1.2-8) [39,40]. Raw data were used to create a vector magnitude √(x^2+ y^2+z^2)-1g negative numbers were rounded to 0 to create the Euclidean Norm minus one (measured in mg), as previously reported [41]. Data were averaged over 5-s epochs. Nonwear was assessed over a 60-minute window, using moving 15-minute increments [42] if the standard deviation of 2 of the 3 axes were less than 13 mg and the value range was less than 50 mg. Participants were mailed the device before randomization and instructed to wear it continuously on their nondominant hand for 7 days from the following morning, without changing their routine PA. To be considered valid for analysis, data were needed for at least 4 days with a minimum of 10 h per day, including at least 1 day on the weekend. Published thresholds were used to determine average daily minutes of activity in light

(LPA), moderate (MPA), vigorous (VPA), and moderate and vigorous (MVPA) intensities [43]. Minutes of activity accumulated in 10-minute bouts were established using an 80% rule, where activity must be sustained above the appropriate threshold for at least 80% of the time [42].

Self-Reported Physical Activity

Minutes per week of MVPA were estimated using Web-based self-completion of the International PA Questionnaire-Short Form (IPAQ-SF) at baseline and at 2-months postrandomization. The IPAQ-SF is a validated measure of PA [44] and has been used in previous behavioral trials promoting PA for depression [29,45] as well as being the most frequently used measure in Web-based studies for PA [46].

Anxiety

The General Anxiety Disorder scale (GAD-7) is a 7-item, 4-point scale (0-3) and was used to assess anxiety using Web-based self-completion at baseline and at 2-months postrandomization. The GAD-7 measures the severity of anxiety symptoms over the past 2 weeks based on the DSM-IV criteria. The GAD-7 has good reliability as well as criterion, construct, factorial, and procedural validity. At the cutoff point of 10, the GAD-7 has a sensitivity of 89% and specificity of 82% [37].

Demographic Data

Data on age, gender, level of education (GCSE, A-levels, degree, postgraduate, or doctoral), employment status (full-time, part-time, homemaker, student, retired, or unemployed), current receipt of psychotherapy (yes or no), current receipt of antidepressants (yes or no), method of recruitment (social media, newspaper, word of mouth, or other), and ethnicity were collected at baseline using a Web-based questionnaire.

Fidelity

As recommended in previous studies [25,47], intervention fidelity was conceptualized and measured in the domains of design fidelity, training fidelity, quality/completeness of delivery, participant receipt, and enactment. The process of establishing good design fidelity for the eMotion intervention was previously reported [36]. However, given that the eMotion intervention had very limited external human support, training fidelity was not applicable in this study. Delivery fidelity was assessed using website usage statistics from the Web-based intervention database. This database provided individual level data about whether the participant registered for eMotion, modules accessed, and the total time spent on each module. Participant receipt and enactment of the intended intervention processes were measured using Web-based questionnaires. For fidelity of receipt, 2 approaches were used. The first approach assessed participants' understanding of how emotions, behaviors, thoughts, and physical feelings affect each other to maintain depression over time. A single item was used based on questions used in a previous study [48]. The item employed a 5-point Likert response scale (Strongly Agree to Strongly Disagree) assessing participant agreement with the following statement: "I understand how emotions, behaviors, thoughts, and physical feelings affect each other to maintain depression over time." The second approach assessed participants' perceived ability to use the intended BCTs. This was assessed by asking participants to rate their confidence in using specific BCTs (ie, identification of suitable activities, grading activities for ease of use, and planning and dealing with setbacks) over the last 2 months on a scale from 1 (not at all confident) to 10 (very confident). This measure was adapted from measures of confidence used in the ProActive trial [49]. Finally, to assess enactment, we asked participants if they had used specific BCTs related to BA in the last 2 months using a binary scale (yes/no). This measure was adapted from similar measures of BCT usage showing that enactment was significantly associated with weight loss, providing initial evidence of the validity of this type of measure [50].

Acceptability

The Client Satisfaction Questionnaire-Short Form (CSQ-SF) is a 4-item measure and was used to assess participant satisfaction regarding their use of eMotion 2 months postrandomization (given to intervention participants only). This measure was administered using a Web-based questionnaire and has been used to assess treatment satisfaction in other studies of Web-based interventions for depression [51].

Sample Size

Due to the pilot nature of the study, no formal sample size calculations were conducted. However, to ensure a suitably reliable estimate of the standard deviations to power a future trial with 90% power, at least 15 people per arm were recommended if the expected effect size was to be between 0.3 and 0.7 [52]. A previous meta-analysis of computer-based psychological treatments for depression reported a moderate effect size (0.56) and a drop-out rate of 57% [38]. As such, a target sample size of 62 was adopted (accounting for a possible attrition rate of 50%) to ensure at least 15 people per arm at follow-up.

Randomization

Once participants completed the baseline assessment, they were randomly allocated to either the intervention or control group using simple randomization at the individual level in a 1:1 ratio and a Web-based randomization service (Sealed Envelope Ltd. 2016). Personal details were anonymized through the use of participant numbers that were entered into the website by the lead author in a consecutive manner (in the order of completed baseline assessment), and the randomization service allocated them to either group A (eMotion) or group B (waiting list) without any stratification.

Blinding

Due to limited resources for the study, the lead author was not blinded to which condition each participant was allocated following randomization. Due to the nature of the intervention, it was also impossible to blind participants to group allocation. However, because outcome measures were taken using Web-based self-report surveys, there was a reduced chance of the lead author influencing the participant's responses or for the lead author to misinterpret responses or introduce subjective bias into recorded observations [53].

Statistical Analysis

Quantitative methods were used to explore the following: recruitment and attrition rates of trial participants; feasibility of data collection and study procedures; baseline data (including levels of PA and baseline differences between groups and between dropouts); and fidelity of delivery. Descriptive statistics were produced for all outcomes by trial arm at baseline and 2-month follow-up. All quantitative analyses were conducted using Stata SE statistical software release 14 (StataCorp. 2015; College Station, TX). No formal hypothesis testing relating to primary outcomes was planned because this was a pilot study. However, descriptive statistics were used to assess recruitment and retention rates and baseline PA levels. Baseline demographic and clinical characteristics were descriptively presented as proportions or as means with standard deviations. Two types of exploratory analyses of the primary outcome (PHQ-8) were conducted: 1) linear regression models to report changes in depression with 95% confidence intervals around the between-group mean difference and 2) logistic regression models that dichotomized the primary outcome to reflect clinically meaningful change (a reduction to below 10 on the PHQ-8 indicated that the person may no longer qualify for major depression) [37]. The analyses were conducted on participants with complete data only, which included those who began treatment and provided follow-up data regardless of treatment compliance. Missing data were not imputed. Similar analyses were conducted for anxiety, objective, and self-reported PA. We conducted sensitivity analyses using linear regression models to examine the effects of receiving psychological therapies as well as any substantial differences in baseline characteristics on the findings. We also analyzed the mean reduction in depressive symptoms for those who received the minimum dose of intervention and provided data at 2 months postrandomization.

Ethics

This study was approved by the University of Exeter Sports and Health Sciences Research Ethics Committee (AM160316-21 151021/B/03). One possible ethical issue in this study was suicide risk in people experiencing depression. Because this was a research study on a nonclinical sample, all participants were advised on the PIS that the study was not a clinical or NHS treatment and that the University and researchers could not take clinical responsibility for the treatment of any conditions they might have including depression. They were also signposted to other appropriate resources in case they wished to seek formal treatment. If, at any point in the study (eg, while on the phone to a researcher during screening or after inclusion), participants indicated suicidal intent, the University of Exeter Mood Disorders Suicide Risk Protocol was invoked.

Results

Participant flow

A total of 183 people responded to the adverts, with 100 completing screening for eligibility (Figure 1). Of the 183 individuals who initially inquired about the study, 100 were still interested and screened for eligibility and 62 (34% of those who initially enquired [95% CI 27-41] and 62% (62/100) of those who were screened [95% CI 52-71]) were eligible for inclusion and randomized in the trial between May 2016 and February 2017 (32 in the eMotion group, 30 in the control group). Overall attrition in relation with the planned main trial primary outcome (PHQ-8) at 2 months postrandomization was 19% (12/62; 95% CI 11-31). Of those randomized, 94% (58/62) of participants provided complete secondary outcome baseline measurements (eg, GAD7, IPAQ-SF) and 84% (52/62) provided usable accelerometer data at baseline. At 2-month follow-up, 81% (50/62) of those randomized [95% CI 71-91] provided PHQ-8 (and other survey data) and 39% (24/62) of those randomized [95% CI 27-52] provided valid accelerometer data. Only 76% (47/62) and 53% (33/62) participants provided valid IPAQ-SF data at baseline and 2 months postrandomization, respectively. This lack of usable IPAQ-SF data was due to people providing invalid responses.

Feasibility of Accelerometer Data Collection

At baseline, 6 accelerometers were not sent out due to participants not responding to the request confirming their willingness to wear it or not being willing /able to wear it. Three accelerometers had data processing problems (technical failure). At follow-up, missing accelerometer data were primarily due to participants not responding to the request to wear the device again (n=13) or participants being lost to follow-up (not responding in any way; n=9).

Baseline Demographic and Clinical Characteristics

At baseline (Table 2 and Table 3), the mean age was 38 years with women accounting for 84% (52/62) of all participants, and 97% (60/62) of participants were white British. Nearly half the sample was recruited through social media (Facebook or Twitter) with the second most popular method being “word of mouth” (ie, hearing about the study from friends or family). Participants had a range of educational levels, and most (55/62, 89%) of them were employed either part-time or full-time. The mean score on the PHQ-8 was 14.6 (SD 3.2), and the mean score on the GAD-7 was 11.8 (SD 4.5). All PA data were positively skewed; hence, medians and interquartile ranges were reported. The median daily total minutes of accelerometer-measured PA was 174.3 (IQR 136.8-212.5) for light PA (LPA), 53.5 (IQR 39.8-80.7) for moderate PA (MPA), 2.9 (IQR 1.0-6.2) for vigorous PA (VPA), and 55.2 (IQR 40.9-90.7) for moderate and vigorous PA (MVPA). The median weekly total minutes of accelerometer-measured MVPA in at least 10-minute bouts was 35.8 (IQR 0.0-98.6). Only 13% (7/52) people achieved at least 150 minutes per week of MVPA in at least 10-minute bouts. The median level of daily self-reported MVPA was 12.9 minutes (IQR 0.0-25.7). Over half (36/62, 58%) of the participants were receiving antidepressants, and 13% (8/62) were receiving some form of psychotherapy, with a higher proportion receiving therapy in the control group (7/30, 23%) than in the intervention group (1/32, 3%). The intervention group had a higher median of total MVPA per day (71 min, IQR 46.7-85.9) than the control group (55 min, IQR 40.1-90.7). Finally, the intervention group was older by 2 years with a mean of 39.3 (12.0) years compared with 36.9 (12.6) years in the control group.

Figure 1. CONSORT flow chart.

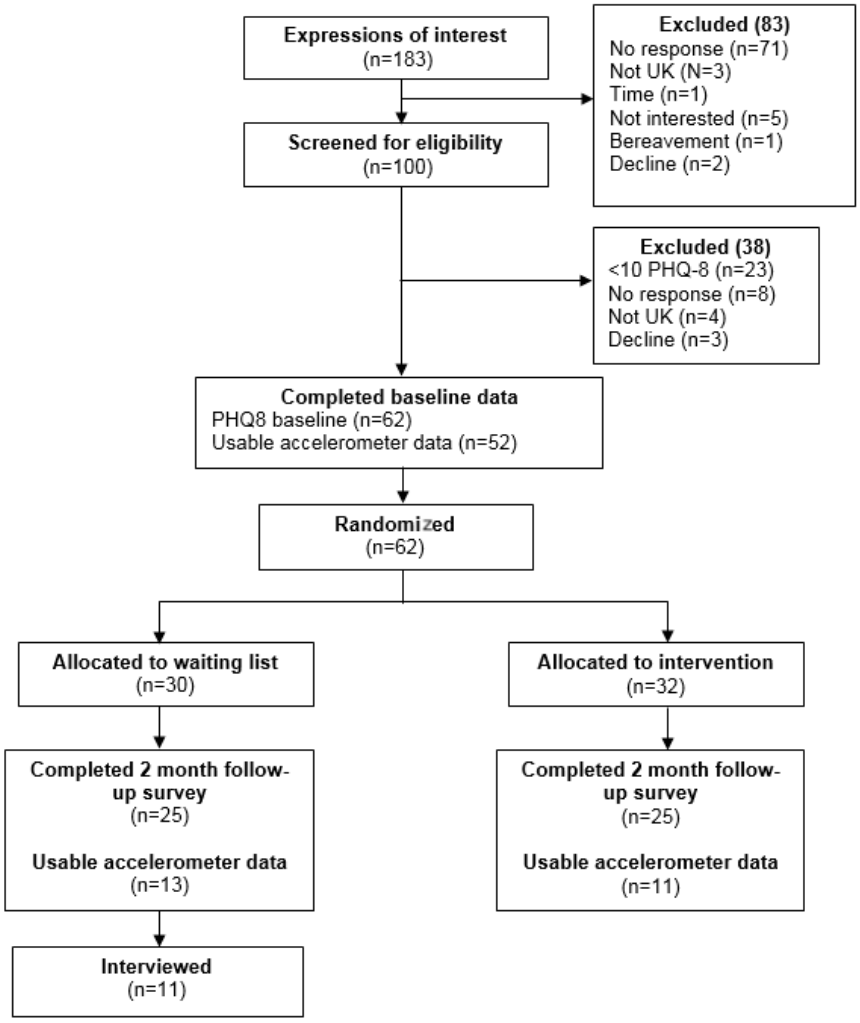


Table 2. Participant demographic and clinical characteristics at baseline.

Characteristics	eMotion		Control group		Whole sample	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Age in years	32	39.3 (12.0)	30	36.9 (12.6)	62	38.1 (12.3)
Depression (PHQ-8[a])	32	14.4 (3.5)	30	14.8 (2.9)	62	14.6 (3.2)
Anxiety (GAD-7[b])	31	11.5 (4.7)	27	12.3 (4.2)	58	11.8 (4.5)
Min per week of objective MVPA[c] in 10-min bouts[d]	27	29.5 (0.0-98.8)	25	42.1 (8.1-93.7)	52	35.8 (0.0-98.6)
IPAQ-SF daily min of MVPA[d]	27	12.9 (0.0-25.7)	20	10.7 (3.6-17.9)	47	12.9 (0.0-25.7)

[a]PHQ-8: Patient Health Questionnaire 8.

[b]GAD-7: General Anxiety Disorder scale.

[c]MVPA: moderate-to-vigorous intensity physical activity.

[d]Data were positively skewed, so medians (interquartile ranges) are reported.

Table 3. Participant demographic and clinical characteristics at baseline.

Characteristics	eMotion		Control Group		Whole Sample	
	N	n (%)	N	n (%)	N	n (%)
Female	32	26 (81)	30	26 (87)	62	52 (84)
Receiving therapy	32	1 (3)	30	7 (23)	62	8 (13)
Antidepressants	32	18 (56)	30	18 (60)	62	36 (58)
>150 m per week of MVPA[a] (10-min bouts)	27	3 (11)	25	4 (16)	52	7 (13)
Currently employed, studying, or training	32	28 (88)	30	27 (90)	62	55 (89)
Educated to A level or beyond	32	25 (78)	30	23 (77)	62	48 (77)

[a]MVPA: moderate-to-vigorous intensity physical activity.

Intervention fidelity

In total, 88%(28/32) of the intervention participants registered for eMotion and began the introduction module. The median number of logins, modules accessed, and total minutes spent on eMotion was 3 (IQR 2-8), 3 (IQR 2-5), and 41.3 (IQR 18.9-90.4), respectively. Overall, 53% (17/32) of participants completed at least the introduction, week 1, and week 2, and 25% (8/32) of participants completed up to at least week 4. Only one participant used every module. Of the 46 participants who provided receipt and enactment data at both baseline and 2 months postrandomization, those randomized to the eMotion group reported a significant difference, compared with the control group, in levels of understanding about how thoughts, feelings, and behaviors affect mood (adjusted mean difference 0.5, 95% CI −0.0 to −1.0). Significant differences were also found for confidence to identify (adjusted mean difference 1.4, 95% CI 0.0-2.8), select (adjusted mean difference 1.3, 95% CI −0.02 to 2.6), and plan (adjusted mean difference 1.8, 95% CI 0.5-3.1) achievable activities to improve mood as well as confidence to deal with setbacks (adjusted mean difference 1.5, 95% CI 0.2-2.7). Of the participants who answered “no” on the enactment questionnaires at baseline, those who were randomized to the eMotion group were significantly more likely to select (N=25; OR 10, 95% CI 1.6-62.7) and plan (N=33; OR 10.3, 95% CI 2.0-52.6) activities to improve their mood at the 2-month follow-up.

Acceptability

Of the participants randomized to receive the eMotion intervention who provided follow-up data, 4% (1/25) felt almost all their needs had been met, 32% (8/25) felt most of their needs had been met, 52% (13/25) felt only a few of their needs had been met, and 12% (3/25) felt none of their needs had been met. Twenty-four percent (6/25) said they would definitely use the program again, 32% (8/25) said “Yes I think so,” 40% (10/25) said “No, I don’t think so,” and 4% (1/25) said “Definitely not.” Finally, 16% (4/25) said they were “Very Satisfied,” 40% (10/25) said they were “Mostly Satisfied,” 40% (10/25) said they were “Indifferent or Mildly Satisfied,” and 4% (1/25) said they were “Quite Dissatisfied.”

Exploratory Analysis of Outcomes

Exploratory analyses carried out on complete data (Table 4) showed that at 2 months postrandomization, the intervention group had a larger reduction in depressive symptoms than the control group (adjusted mean difference −3.6, 95% CI −6.1 to −1.1). In the intervention group, 56% (14/25) of depression scores went below the threshold of 10 on the PHQ-8, compared with 28% (7/25) in the control group (OR 3.3, 95% CI 1.0-10.6). For those who completed the minimum dose of intervention and provided data at 2 months postrandomization (n=15), the mean reduction in depressive symptoms was 5 points (SD 5.4). For those who did not complete the minimum dose of intervention and provided data at 2 months postrandomization (n=9), the mean reduction in depressive symptoms was 4.9 points (SD 4.6). Of the 47 participants who provided anxiety scores at both baseline and 2 months postrandomization, there was a larger reduction in symptoms of anxiety for the eMotion group than the control group (adjusted mean difference −3.3, 95% CI −5.4 to −1.2). Linear regression analysis on complete data, controlling for baseline PA, revealed no between-group differences in PA at any intensity. Valid IPAQ-SF data were available for 33 trial participants at 2 months postrandomization. Linear regression analysis controlling for baseline PA revealed no between-group differences in self-reported PA.

Sensitivity analysis

When receipt of other psychological therapies was entered into the regression analysis as a covariate, the impact of co-treatment was not significant, and the intervention group still had a higher reduction in depressive symptoms than the control group (adjusted mean difference −3.3, 95% CI −5.9 to −0.7). Other baseline covariates (age, gender, employment, education level, and antidepressant usage) that may have influenced depression scores were also entered in the regression model together. Findings indicated that none of these variables had a significant covariate effect on depression scores and that the residual difference between groups was still significant (adjusted mean difference −3.1, 95% CI −5.7 to −0.5). Within the intervention group, linear regression analyses revealed no significant relationships between numbers of modules accessed, number of logins, or total minutes spent on the website with depression outcomes. The pattern of change scores within each group is shown in Figure 2.

Table 4. Between-group changes in primary and secondary outcomes.

Outcomes	eMotion		Control group		Adjusted mean difference[a] (95% CI)
	N	Mean (SD)	N	Mean (SD)	
Depression (PHQ-8[b])					
Baseline	32	14.4 (3.4)	30	14.8 (2.9)	
2 months postrandomization	25	8.7 (4.8)	25	12.9 (4.2)	−3.6 (−6.1 to −1.1)
Anxiety (GAD-7[c])					
Baseline	31	10.1 (5.4)	27	12.0 (4.7)	
2 months postrandomization	25	7.1 (3.8)	25	10.9 (3.7)	−3.3 (−5.4 to −1.2)
Min per week of objective MVPA in 10-min bouts[d]					
Baseline	27	29.5 (0.0 to 8.8)	25	42.1 (8.1 to 93.7)	
2 months postrandomization	13	97.6 (49.7 to 166.3)	11	13.0 (0.0 to 131.4)	16.4 (−43.7 to 76.5)
IPAQ-SF daily minutes of MVPA[d]					
Baseline	27	12.9 (0.0 to 25.7)	20	10.7 (3.6 to 17.9)	
2 months postrandomization	19	11.4 (4.3 to 25.7)	14	15.7 (0.0 to 22.9)	0.2 (−8.7 to 9.2)

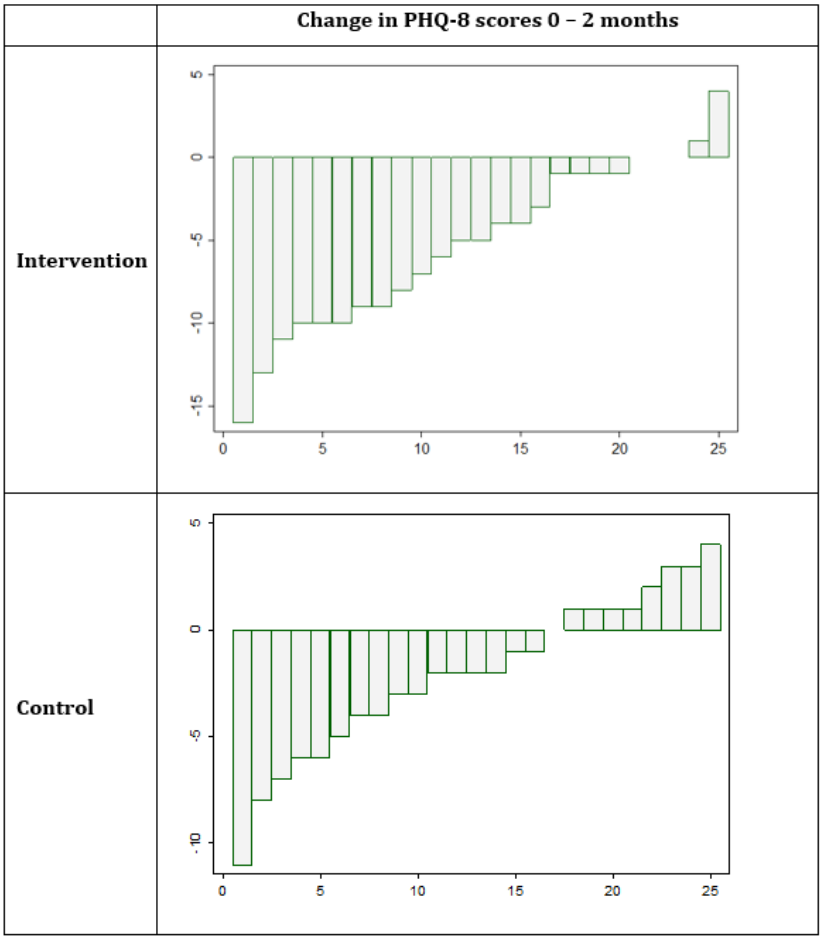
[a]Multiple regression adjusted for baseline value and confidence intervals reported.

[b]PHQ-8: Patient Health Questionnaire 8.

[c]GAD-7: General Anxiety Disorder scale.

[d]As physical activity data were skewed, medians and interquartile ranges (IQR) are presented, and analysis was repeated using bootstrapping.

Figure 2. PHQ-8 change scores for individual participants.



Discussion

Summary of Findings

The present study examined the feasibility of conducting an RCT of the eMotion intervention. We successfully recruited a less-active population with elevated depression and anxiety. The trial also had acceptable attrition rates concerning the primary outcome at 2 months. Concerning the feasibility of data collection, most people provided valid accelerometer data at baseline. However, there was a lack of valid accelerometer data at 2 months postrandomization. The baseline level of MVPA in 10-minute bouts was low, with only 13% (7/52) of participants achieving at least 150 minutes of MVPA per week. Of those randomized to the intervention group, most people registered for eMotion and just over half completed the "minimum dose" (introduction, week 1, and week 2). Exploratory analyses revealed larger changes in depressive symptoms and anxiety in the eMotion group than those in the control group at 2 months postrandomization. Descriptive PA data revealed a higher weekly median of minutes of MVPA in 10-minute bouts per week in the eMotion group than in the control group at 2 months postrandomization. This difference was not significant, although this may reflect low numbers available for analysis and high variance in this measure.

Relationship to Other Literature

The achieved recruitment rates resembled those found in other studies of Web-based interventions promoting PA for depression [29-31,45,54]. Other studies report between 26% and 46% of people approached being subsequently randomized [29-31,45,54]. Mailey et al [54] recruited students registered with mental health counseling services but did not have an inclusion criterion for the level of depressive symptoms, possibly contributing to their higher recruitment rate (46%).

Our retention rate compared well with other self-delivered trials of psychological and PA interventions for depression [30]. A recent 3-arm RCT comparing administratively supported Web-based BA, PA, and a waiting list control group had a similar retention rate to ours at 8 weeks postrandomization (82%) [30]. Conversely, a systematic review of Web-based psychological treatments for depression reported drop-out rates of 74% for unsupported, 38.4% for administratively supported, and 28% for therapist-supported treatments [38]. Some individual trials of Web-based PA interventions for depression have reported lower attrition rates than ours (8%-12%) [29,54]. However, both of these trials provided therapist support, whereas our study only provided administrative support. Our data, combined with previous research, suggested that a low level of administrative support may be adequate to retain acceptable numbers of participants at follow-up.

Most studies investigating the effect of PA on depression conceptualize PA as a prescribed structured intervention and have not previously measured baseline levels or changes in PA [6,8]. Also, change in PA has typically not been measured or reported in trials of psychological treatments for depression [55]. This is one of the few intervention studies to collect baseline PA data (using objective measures) in people with elevated depressive symptoms, building on other recent intervention studies following a similar approach [56].

With a baseline median of 35.8 (IQR 0.0-98.6) minutes per week of MVPA (in at least 10-minute bouts), we appear to have recruited a less-active sample than other similar studies. For example, a recent cross-sectional study (n=165) of adults with depression (≥10 on the PHQ-9) reported a baseline mean of 18.2 (SD 17.4) minutes of MVPA (in 10-minute bouts) per day [57]. One possible explanation for this is that the clear indication of "exercise" in the PIS attracted people with depression who were already active [57]. In eMotion, however, the intervention was not overtly presented as exercise, but rather as a behavioral intervention to promote routine, pleasurable, and necessary activities (which could include PA).

The current guidelines for PA (150 minutes of MVPA per week) are based on improving and maintaining physical health, rather than mental health [58]. The dose of PA for improving and maintaining mental health is not clear and may well be linked to the quality of the experience rather than just the physical volume. For example, a recent longitudinal cohort study (n=33,908) suggests that 12% of future cases of depression could be prevented with just 60 minutes of any intensity PA per week [59].

Only 56% of the participants in the intervention arm accessed at least the introduction, week 1, and week 2 (minimum dose), suggesting that more could be done to draw potential users into the website. eMotion actively encouraged participants to engage with the process of BA in their day-to-day lives (eg, planning and reviewing goals using their diaries). This was reflected in our process measures of receipt and enactment showing that despite the relatively low usage statistics, people randomized to eMotion were more confident to identify, select, and plan activities to improve their mood as well as to deal with setbacks (ie, to engage in the key processes of behavior change proposed by the eMotion logic model) [36]. This explanation is consistent with a recent observational study (n=8993) of a Web-based handwashing intervention (PRIMIT) [60]. In PRIMIT, the largest change in behavior occurred after the first session, with incrementally smaller changes occurring after each subsequent session [60]. Taken together, these findings suggest that usage metrics reveal little about offline engagement with intervention processes and that usage cessation could either indicate disengagement from the intervention or the development of sufficient mastery [61].

Exploratory analysis revealed a decrease in symptoms of depression and anxiety in favor of the intervention group at 2 months postrandomization, a similar reduction to that found in previous studies promoting PA for depression [23,29-31,45] and anxiety [13]. However, other studies did not find such an effect [54,62,63], possibly due to low power and the use of active control conditions [7]. Our findings tentatively support the utility of using BA as a more-general treatment for depression and anxiety and are consistent with findings from a large-scale RCT that found BA to be no less effective than cognitive behavioral therapy for treating depression [64]. However, it is important to note that due to its low power,

definitive conclusions around effectiveness cannot be made from the current study.

Strengths and Limitations

This is the first study to evaluate the feasibility of delivering BA in combination with the explicit promotion of PA, in a Web-based format. The main strength of this study was the use of rigorous methods to assess the feasibility of conducting a full-scale RCT. We used objective methods to assess PA and validated self-report measures of depression and anxiety symptoms. However, several limitations of this study need to be acknowledged.

As observed in other large-scale depression trials, our sample did not represent the wider UK population, particularly in terms of ethnicity and gender [64,65]. This is most likely an artifact of our recruitment method (ie, community recruitment) and target location (South West England). However, there are other individual factors including ability to recognize and accept mental health problems, positive impact of social networks, reluctance to discuss psychological distress and seek help among men, cultural identity, perceived social stigma against mental health, and financial factors [66]. A larger trial could attempt to recruit a more representative sample by targeting locations with more culturally diverse populations. Tailored recruitment approaches could also be used to address individual barriers to engagement (eg, using adverts targeted at males).

Although we randomized participants, due to a lack of resources for independent data collection, there was no blinding, which could have led to an inflation of the observed effects [67]. However, the potential for researcher bias was limited in this case due to an absence of any face-to-face contact when collecting outcome measures. Our groups were imbalanced at baseline with regard to co-interventions. However, these factors did not seem to strongly impact the findings. A future trial could remedy this either by including therapy as a randomization-minimization variable or adding it as an exclusion criterion.

Due to resource issues, "minimal contact" administrative support was provided by the lead author. It is possible that this may have led to bias when collecting outcomes. However, because the outcomes were collected via self-administered Web-based questionnaires and accelerometers, this is unlikely to have had an effect. We would still recommend using independent supporters in the main trial for practical reasons and for testing the feasibility of providing such support in a "real-world" NHS context, if the intervention proves effective in a full-scale RCT.

A further limitation is that the PHQ-8 was used rather than the more conventional PHQ-9. The PHQ-8 was chosen due to the lack of any directly available clinical surveillance or support for participants, as it would not have been feasible to follow-up any (Web-based) survey responses expressing suicidal ideation in response to PHQ item 9 with an immediate telephone interview. The PHQ-8 is specifically recommended for use in such circumstances [37]. Furthermore, the PHQ-8 is very similar to the PHQ-9 and has excellent convergent validity (r=0.997), indicating that the 2 scales are comparable [68].

Implications for Future Research

Future studies should refine procedures (as indicated above) and further develop the eMotion intervention to optimize user engagement and experiences. Despite exploratory data showing modest reductions in depression and anxiety, only half of the people who used eMotion were mostly or very satisfied with their experience. Qualitative interviews performed on a sample of participants (n=11) have helped to identify barriers and facilitators to engaging with the intervention and with the trial (including the use of accelerometers) and suggest ways to maximize data collection and minimize attrition. This data will be reported in detail elsewhere. Refinements of the study procedures would also be needed to collect more complete and meaningful data on PA in any future trial. This could be achieved via face-to-face contact or by providing incentives.

In line with the MRC framework [69], large, well-controlled RCTs that build on the findings from this pilot trial could help to more definitively test whether such an intervention is effective in reducing depression and increasing PA in community-dwelling populations with depression in the UK and elsewhere.

Conclusion

The eMotion intervention is novel in attempting to offer an integrated solution to the 2 critical public health priorities of depression and lack of PA. Based on the data presented, both the eMotion intervention and methods needed to conduct a trial seem to be feasible and acceptable. If successful in a large-scale trial, eMotion would have the potential to reduce depression and anxiety symptoms for people in the community, easing the burden on NHS resources. There may also be further potential to increase PA in this population.

Acknowledgments

JDL's time input was supported by the Economic and Social Research Council (ESRC Grant Number: ES/J50015X/1). This report is independent research and the views expressed are those of the authors and not necessarily those of National Institute for Health Research (NIHR) or the UK Department of Health. At the time of manuscript submission, AHT is funded by the NIHR Health Technology Assessment Program (ref: 15/111/01; ref 13/25/20), the NIHR Public Health Research Program (14/54/19), the NIHR Research for Patient Benefit Program, and the Medical Research Council, intervention development program.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1MB - jmir_v20i7e10112_app1.pdf]

References

1. Stevenson D, Farmer P. GOV.UK. Thriving at work: The Stevenson / Farmer review of mental health and employers URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/658145/thriving-at-work-stevenson-farmer-review.pdf [WebCite Cache ID 6zlpg6X94]
2. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The prevalence of co-morbid depression in adults with Type 2 diabetes: a systematic review and meta-analysis. Diabet Med 2006 Nov;23(11):1165-1173. [doi: 10.1111/j.1464-5491.2006.01943.x] [Medline: 17054590]
3. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. Diabetes Care 2001 Jun;24(6):1069-1078. [Medline: 11375373]
4. Whooley MA, de JP, Vittinghoff E, Otte C, Moos R, Carney RM, et al. Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease. JAMA 2008 Nov 26;300(20):2379-2388 [FREE Full text] [doi: 10.1001/jama.2008.711] [Medline: 19033588]
5. Onyike CU, Crum RM, Lee HB, Lyketsos CG, Eaton WW. Is obesity associated with major depression? Results from the Third National Health and Nutrition Examination Survey. Am J Epidemiol 2003 Dec 15;158(12):1139-1147. [Medline: 14652298]
6. Cooney GM, Dwan K, Greig CA, Lawlor DA, Rimer J, Waugh FR, et al. Exercise for depression. Cochrane Database Syst Rev 2013;9:CD004366. [doi: 10.1002/14651858.CD004366.pub6] [Medline: 24026850]
7. Ekkekakis P. Honey, I shrunk the pooled SMD! Guide to critical appraisal of systematic reviews and meta-analyses using the Cochrane review on exercise for depression as example. Mental Health and Physical Activity 2015 Mar;8:21-36. [doi: 10.1016/j.mhpa.2014.12.001]
8. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev 2015;9(3):366-378. [doi: 10.1080/17437199.2015.1022901] [Medline: 25739893]
9. Mammen G, Faulkner G. Physical activity and the prevention of depression: a systematic review of prospective studies. Am J Prev Med 2013 Nov;45(5):649-657. [doi: 10.1016/j.amepre.2013.08.001] [Medline: 24139780]
10. Searle A, Calnan M, Lewis G, Campbell J, Taylor A, Turner K. Patients' views of physical activity as treatment for depression: a qualitative study. Br J Gen Pract 2011 Apr;61(585):149-156 [FREE Full text] [doi: 10.3399/bjgp11X567054] [Medline: 21439172]
11. Ussher M, Stanbury L, Cheeseman V, Faulkner G. Physical activity preferences and perceived barriers to activity among persons with severe mental illness in the United Kingdom. Psychiatr Serv 2007 Mar;58(3):405-408. [doi: 10.1176/ps.2007.58.3.405] [Medline: 17325117]
12. Babyak M, Blumenthal JA, Herman S, Khatri P, Doraiswamy M, Moore K, et al. Exercise treatment for major depression: maintenance of therapeutic benefit at 10 months. Psychosom Med 2000;62(5):633-638. [Medline: 11020092]
13. Stubbs B, Vancampfort D, Rosenbaum S, Firth J, Cosco T, Veronese N, et al. An examination of the anxiolytic effects of exercise for people with anxiety and stress-related disorders: A meta-analysis. Psychiatry Res 2017 Dec;249:102-108. [doi: 10.1016/j.psychres.2016.12.020] [Medline: 28088704]
14. Gibala MJ, Little JP, Macdonald MJ, Hawley JA. Physiological adaptations to low-volume, high-intensity interval training in health and disease. J Physiol 2012 Mar 01;590(5):1077-1084 [FREE Full text] [doi: 10.1113/jphysiol.2011.224725] [Medline: 22289907]
15. Faulkner G, Taylor A. Promoting physical activity for mental health: A complex intervention? Mental Health and Physical Activity 2009 Jun;2(1):1-3. [doi: 10.1016/j.mhpa.2009.04.001]
16. Rebar AL, Taylor A. Physical activity and mental health; it is more than just a prescription. Mental Health and Physical Activity 2017 Oct;13:77-82. [doi: 10.1016/j.mhpa.2017.10.004]
17. Hopko DR, Lejuez CW, Ruggiero KJ, Eifert GH. Contemporary behavioral activation treatments for depression: procedures, principles, and progress. Clin Psychol Rev 2003 Oct;23(5):699-717. [Medline: 12971906]
18. Lejuez CW, Hopko DR, Hopko SD. A brief behavioral activation treatment for depression. Treatment manual. Behav Modif 2001 Apr;25(2):255-286. [doi: 10.1177/0145445501252005] [Medline: 11317637]
19. Richards D. Behavioural Activation. In: Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh D, et al, editors. Oxford guide to low intensity CBT interventions. Oxford: Oxford University Press; 2010:141-150.
20. Ekers D, Webster L, Van SA, Cuijpers P, Richards D, Gilbody S. Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. PLoS One 2014;9(6):e100100 [FREE Full text] [doi: 10.1371/journal.pone.0100100] [Medline: 24936656]

21. 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]
22. Farrand P, Pentecost C, Greaves C, Taylor RS, Warren F, Green C, et al. A written self-help intervention for depressed adults comparing behavioural activation combined with physical activity promotion with a self-help intervention based upon behavioural activation alone: study protocol for a parallel group pilot randomised controlled trial (BAcPAc). Trials 2014 May 29;15:196 [FREE Full text] [doi: 10.1186/1745-6215-15-196] [Medline: 24886116]
23. Pentecost C, Farrand P, Greaves CJ, Taylor RS, Warren FC, Hillsdon M, et al. Combining behavioural activation with physical activity promotion for adults with depression: findings of a parallel-group pilot randomised controlled trial (BAcPAc). Trials 2015 Aug 20;16:367 [FREE Full text] [doi: 10.1186/s13063-015-0881-0] [Medline: 26289425]
24. Watkins E, Newbold A, Tester-Jones M, Javaid M, Cadman J, Collins LM, et al. Implementing multifactorial psychotherapy research in online virtual environments (IMPROVE-2): study protocol for a phase III trial of the MOST randomized component selection method for internet cognitive-behavioural therapy for depression. BMC Psychiatry 2016 Dec 06;16(1):345 [FREE Full text] [doi: 10.1186/s12888-016-1054-8] [Medline: 27716200]
25. Lambert JD, Greaves CJ, Farrand P, Cross R, Haase AM, Taylor AH. Assessment of fidelity in individual level behaviour change interventions promoting physical activity among adults: a systematic review. BMC Public Health 2017 Dec 02;17(1):765 [FREE Full text] [doi: 10.1186/s12889-017-4778-6] [Medline: 28969669]
26. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
27. Farmer C, Farrand P, O'Mahen H. 'I am not a depressed person': how identity conflict affects help-seeking rates for major depressive disorder. BMC Psychiatry 2012 Oct 02;12:164 [FREE Full text] [doi: 10.1186/1471-244X-12-164] [Medline: 23031330]
28. Carlbring P, Lindner P, Martell C, Hassmén P, Forsberg L, Ström L, et al. The effects on depression of Internet-administered behavioural activation and physical exercise with treatment rationale and relapse prevention: study protocol for a randomised controlled trial. Trials 2013 Feb 02;14:35 [FREE Full text] [doi: 10.1186/1745-6215-14-35] [Medline: 23374879]
29. Nyström MBT, Stenling A, Sjöström E, Neely G, Lindner P, Hassmén P, et al. Behavioral activation versus physical activity via the internet: A randomized controlled trial. J Affect Disord 2017 Dec;215:85-93 [FREE Full text] [doi: 10.1016/j.jad.2017.03.018] [Medline: 28319696]
30. Soucy I, Provencher M, Fortier M, McFadden T. Efficacy of guided self-help behavioural activation and physical activity for depression: a randomized controlled trial. Cogn Behav Ther 2017 Nov;46(6):493-506. [doi: 10.1080/16506073.2017.1337806] [Medline: 28644740]
31. Rosenbaum S, Newby JM, Steel Z, Andrews G, Ward PB. Online physical activity interventions for mental disorders: A systematic review. Internet Interventions 2015 May;2(2):214-220. [doi: 10.1016/j.invent.2015.04.001]
32. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
33. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687 [FREE Full text] [Medline: 24609605]
34. Farrand P, Woodford J. Impact of support on the effectiveness of written cognitive behavioural self-help: a systematic review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2013 Feb;33(1):182-195. [doi: 10.1016/j.cpr.2012.11.001] [Medline: 23238023]
35. van GJEWC, Nijland N, van LM, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
36. Lambert JD, Greaves CJ, Farrand P, Haase AM, Taylor AH. Development of a web-based intervention (eMotion) based on behavioural activation to promote physical activity in people with depression. Mental Health and Physical Activity 2017 Oct;13:120-136. [doi: 10.1016/j.mhpa.2017.10.003]
37. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
38. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
39. van HVT, Gorzelniak L, Dean LEC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 2013;8(4):e61691 [FREE Full text] [doi: 10.1371/journal.pone.0061691] [Medline: 23626718]
40. Sabia S, van HVT, Shipley MJ, Trenell MI, Hagger-Johnson G, Elbaz A, et al. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol 2014 Mar 15;179(6):781-790 [FREE Full text] [doi: 10.1093/aje/kwt330] [Medline: 24500862]

41. Hildebrand M, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc 2014 Sep;46(9):1816-1824. [doi: 10.1249/MSS.0000000000000289] [Medline: 24887173]
42. da Silva IC, van Hees VT, Ramires VV, Knuth AG, Bielemann RM, Ekelund U, et al. Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry. Int J Epidemiol 2014 Dec;43(6):1959-1968 [FREE Full text] [doi: 10.1093/ije/dyu203] [Medline: 25361583]
43. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. Med Sci Sports Exerc 2011 Jun;43(6):1085-1093. [doi: 10.1249/MSS.0b013e31820513be] [Medline: 21088628]
44. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
45. Ström M, Uckelstam C, Andersson G, Hassmén P, Umefjord G, Carlbring P. Internet-delivered therapist-guided physical activity for mild to moderate depression: a randomized controlled trial. PeerJ 2013;1:e178 [FREE Full text] [doi: 10.7717/peerj.178] [Medline: 24109561]
46. Joseph RP, Durant NH, Benitez TJ, Pekmezi DW. Internet-Based Physical Activity Interventions. Am J Lifestyle Med 2014 Dec;8(1):42-68 [FREE Full text] [doi: 10.1177/1559827613498059] [Medline: 25045343]
47. Borrelli B. The Assessment, Monitoring, and Enhancement of Treatment Fidelity In Public Health Clinical Trials. J Public Health Dent 2011;71(s1):S52-S63 [FREE Full text] [doi: 10.1111/j.1752-7325.2011.00233.x] [Medline: 21499543]
48. Brawley LR, Arbour-Nicitopoulos KP, Martin GKA. Developing physical activity interventions for adults with spinal cord injury. Part 3: a pilot feasibility study of an intervention to increase self-managed physical activity. Rehabil Psychol 2013 Aug;58(3):316-321. [doi: 10.1037/a0032814] [Medline: 23978087]
49. Hardeman W, Michie S, Fanshawe T, Prevost AT, Mcloughlin K, Kinmonth AL. Fidelity of delivery of a physical activity intervention: predictors and consequences. Psychol Health 2008;23(1):11-24. [doi: 10.1080/08870440701615948] [Medline: 25159904]
50. Hankonen N, Sutton S, Prevost AT, Simmons RK, Griffin SJ, Kinmonth AL, et al. Which behavior change techniques are associated with changes in physical activity, diet and body mass index in people with recently diagnosed diabetes? Ann Behav Med 2015 Feb;49(1):7-17 [FREE Full text] [doi: 10.1007/s12160-014-9624-9] [Medline: 24806469]
51. Williams C, Wilson P, Morrison J, McMahon A, Walker A, Andrew W, et al. Guided self-help cognitive behavioural therapy for depression in primary care: a randomised controlled trial. PLoS One 2013;8(1):e52735 [FREE Full text] [doi: 10.1371/journal.pone.0052735] [Medline: 23326352]
52. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
53. Edwards P. Questionnaires in clinical trials: guidelines for optimal design and administration. Trials 2010;11:2 [FREE Full text] [doi: 10.1186/1745-6215-11-2] [Medline: 20064225]
54. 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]
55. Cuijpers P, de Wit L, Taylor A. The effects of psychological treatments for adult depression on physical activity: A systematic review. Mental Health and Physical Activity 2014 Mar;7(1):6-8. [doi: 10.1016/j.mhpa.2014.01.002]
56. Helgadóttir B, Owen N, Dunstan DW, Ekblom Ö, Hallgren M, Forsell Y. Changes in physical activity and sedentary behavior associated with an exercise intervention in depressed adults. Psychology of Sport and Exercise 2017 May;30:10-18. [doi: 10.1016/j.psychsport.2017.01.006]
57. Helgadóttir B, Forsell Y, Ekblom Ö. Physical activity patterns of people affected by depressive and anxiety disorders as measured by accelerometers: a cross-sectional study. PLoS One 2015;10(1):e0115894 [FREE Full text] [doi: 10.1371/journal.pone.0115894] [Medline: 25585123]
58. GOV.UK. Physical activity guidelines for Adults (19–64 years) URL: https://www.gov.uk/government/publications/uk-physical-activity-guidelines [accessed 2018-04-25] [WebCite Cache ID 6ywMz0wvb]
59. Harvey SB, Øverland S, Hatch SL, Wessely S, Mykletun A, Hotopf M. Exercise and the Prevention of Depression: Results of the HUNT Cohort Study. Am J Psychiatry 2018 Dec 01;175(1):28-36. [doi: 10.1176/appi.ajp.2017.16111223] [Medline: 28969440]
60. Ainsworth B, Steele M, Stuart B, Joseph J, Miller S, Morrison L, et al. Using an Analysis of Behavior Change to Inform Effective Digital Intervention Design: How Did the PRIMIT Website Change Hand Hygiene Behavior Across 8993 Users? Ann Behav Med 2017 Jun;51(3):423-431 [FREE Full text] [doi: 10.1007/s12160-016-9866-9] [Medline: 27909944]
61. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
62. Chalder M, Wiles NJ, Campbell J, Hollinghurst SP, Searle A, Haase AM, et al. A pragmatic randomised controlled trial to evaluate the cost-effectiveness of a physical activity intervention as a treatment for depression: the treating depression with

physical activity (TREAD) trial. Health Technol Assess 2012;16(10):1-164, iii [FREE Full text] [doi: 10.3310/hta16100] [Medline: 22398106]
63. Vickers KS, Patten CA, Lewis BA, Clark MM, Ussher M, Ebbert JO, et al. Feasibility of an exercise counseling intervention for depressed women smokers. Nicotine Tob Res 2009 Aug;11(8):985-995 [FREE Full text] [doi: 10.1093/ntr/ntp101] [Medline: 19541948]
64. Richards DA, Ekers D, McMillan D, Taylor RS, Byford S, Warren FC, et al. Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial. Lancet 2016 Aug 27;388(10047):871-880 [FREE Full text] [doi: 10.1016/S0140-6736(16)31140-0] [Medline: 27461440]
65. Wiles N, Thomas L, Abel A, Barnes M, Carroll F, Ridgway N, et al. Clinical effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: the CoBalT randomised controlled trial. Health Technol Assess 2014 May;18(31):1-167, vii. [doi: 10.3310/hta18310] [Medline: 24824481]
66. Memon A, Taylor K, Mohebati LM, Sundin J, Cooper M, Scanlon T, et al. Perceived barriers to accessing mental health services among black and minority ethnic (BME) communities: a qualitative study in Southeast England. BMJ Open 2016 Dec 16;6(11):e012337 [FREE Full text] [doi: 10.1136/bmjopen-2016-012337] [Medline: 27852712]
67. Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. J Hum Reprod Sci 2011 Jan;4(1):8-11 [FREE Full text] [doi: 10.4103/0974-1208.82352] [Medline: 21772732]
68. Razykov I, Ziegelstein RC, Whooley MA, Thombs BD. The PHQ-9 versus the PHQ-8--is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the Heart and Soul Study. J Psychosom Res 2012 Sep;73(3):163-168. [doi: 10.1016/j.jpsychores.2012.06.001] [Medline: 22850254]
69. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [Medline: 18824488]

Abbreviations

BA: behavioral activation
BCTs: behavior change techniques
CeHReS: Centre for eHealth Research and Disease Management
CSQ-SF: Client Satisfaction Questionnaire-Short Form
GAD-7: General Anxiety Disorder scale
IPAQ-SF: International PA Questionnaire-Short Form
MVPA: moderate-to-vigorous intensity physical activity
PA: physical activity
PHQ-8: Patient Health Questionnaire 8
PIS: participant information sheet
RCT: randomized controlled trial
SMD: standardized mean difference
TIDieR: Template for Intervention Description and Replication

Edited by R Calvo, M Czerwinski, J Torous, G Wadley; submitted 13.02.18; peer-reviewed by M Stuckey, E Dove, K Ng; comments to author 11.04.18; revised version received 08.05.18; accepted 10.05.18; published 16.07.18.

Please cite as:
Lambert JD, Greaves CJ, Farrand P, Price L, Haase AM, Taylor AH
Web-Based Intervention Using Behavioral Activation and Physical Activity for Adults With Depression (The eMotion Study): Pilot Randomized Controlled Trial
J Med Internet Res 2018;20(7):e10112
URL: http://www.jmir.org/2018/7/e10112/
doi:10.2196/10112
PMID:30012547

Original Paper

Exploring User Needs for a Mobile Behavioral-Sensing Technology for Depression Management: Qualitative Study

Jingbo Meng[1], PhD; Syed Ali Hussain[1], MA; David C Mohr[2], PhD; Mary Czerwinski[3], PhD; Mi Zhang[4], PhD

[1]Department of Communication, Michigan State University, East Lansing, MI, United States

[2]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Microsoft Research, Redmond, WA, United States

[4]Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, United States

Corresponding Author:
Jingbo Meng, PhD
Department of Communication
Michigan State University
East Lansing, MI,
United States
Phone: 1 5173553480
Email: jingbome@msu.edu

Abstract

Background: Today, college students are dealing with depression at some of the highest rates in decades. As the primary mental health service provider, university counseling centers are limited in their capacity and efficiency to provide mental health care due to time constraints and reliance on students' self-reports. A mobile behavioral-sensing platform may serve as a solution to enhance the efficiency and accessibility of university counseling services.

Objective: The main objectives of this study are to (1) understand the usefulness of a mobile sensing platform (ie, iSee) in improving counseling services and assisting students' self-management of their depression conditions, and (2) explore what types of behavioral targets (ie, meaningful information extracted from raw sensor data) and feedback to deliver from both clinician and students' perspectives.

Methods: We conducted semistructured interviews with 9 clinicians and 12 students with depression recruited from a counseling center at a large Midwestern university. The interviews were 40-50 minutes long and were audio recorded and transcribed. The interview data were analyzed using thematic analysis with an inductive approach. Clinician and student interviews were analyzed separately for comparison. The process of extracting themes involved iterative coding, memo writing, theme revisits, and refinement.

Results: From the clinician perspective, the mobile sensing platform helps to improve counseling service by providing objective evidence for clinicians and filling gaps in clinician-patient communication. Clinicians suggested providing students with their sensed behavioral targets organized around personalized goals. Clinicians also recommended delivering therapeutic feedback to students based on their sensed behavioral targets, including positive reinforcement, reflection reminders, and challenging negative thoughts. From the student perspective, the mobile sensing platform helps to ease continued self-tracking practices. Students expressed their need for integrated behavioral targets to understand correlations between behaviors and depression. They also pointed out that they would prefer to avoid seeing negative feedback.

Conclusions: Although clinician and student participants shared views on the advantages of iSee in supporting university counseling, they had divergent opinions on the types of behavioral targets and feedback to be provided via iSee. This exploratory work gained initial insights into the design of a mobile sensing platform for depression management and informed a more conclusive research project for the future.

(J Med Internet Res 2018;20(7):e10139) doi:10.2196/10139

KEYWORDS

mobile sensing; mental health; depression; counseling; user-centered design

Introduction

Today, college students are dealing with depression at some of the highest rates in decades. According to the 2017 National College Health Assessment, more than one-third of students had "felt so depressed that it was difficult to function" and more than two-thirds had "felt hopeless" within the previous school year [1]. Depression is associated with many other significant problems facing college students including alcohol and substance abuse, eating disorders, dropout, self-injury, and suicide [2]. Responding to this mental health issue is imperative on college campuses.

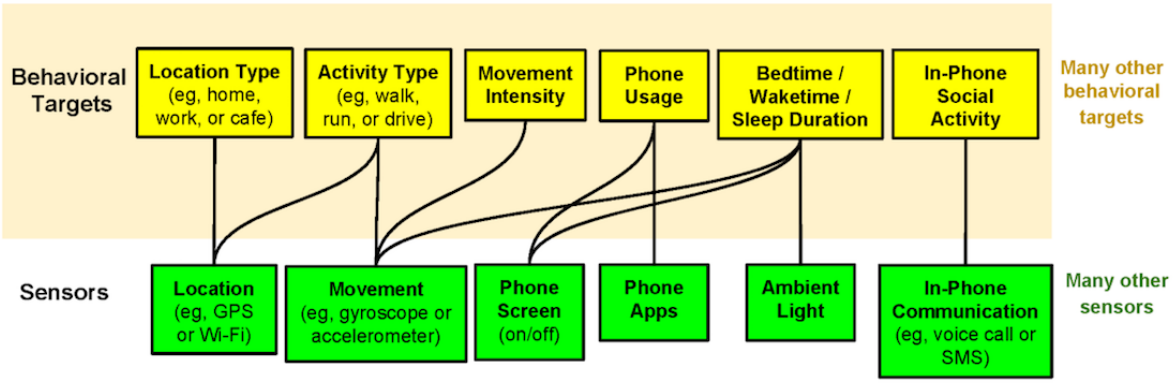
University counseling centers are the primary source for students to access mental health care on college campuses. Unfortunately, there is a lack of capacity and efficiency for university counseling centers to provide counseling services to all students whenever they need. With regard to capacity, the demand of students with depression concerns have been rapidly increasing [3], outpacing the number of staff and working hours that counseling centers could provide. For efficiency, clinicians' reliance on students' self-reports may delay accurate depression assessment and effective treatment delivery. Students with depression are likely to miss clinical appointments [4] or are unable to provide clear or complete information for a variety of reasons, for example forgetfulness or embarrassment [5]. When patients' self-reports are unavailable, or not able to provide an accurate and complete profile about the patients, clinicians may have a difficult time assessing patients' depression conditions, monitoring therapy outcomes [6], and delivering effective therapies.

We propose a large project that aims to design, develop, and implement a mobile sensing platform to address the current challenges that university counseling centers are facing. As the first step of the large project, the present study focuses on gaining insights into the usefulness and design of a mobile sensing platform in enhancing the counseling services available. A mobile sensing platform consists of three components. First, the platform relies on a variety of sensors on mobile and wearable devices which detect and measure physical properties of humans and their environment [7]. For example, smartwatches and smartphones contain onboard sensors that track people's location, movement, sleep, and communication, as well as light and sound in the environment. Second, the platform converts the raw sensor data into behavioral targets through data analytics algorithms [7]. Behavioral targets are meaningful constructs measured by the raw sensor data. For example, raw sensor data may detect ambient light, sound, body movement, and whether the phone screen is on or off and potential behavioral targets can be extracted with regard to bedtime or waketime and sleep duration. These behavioral targets can serve as indicators of depressive symptoms (see Figure 1). Third, the platform delivers behavioral targets and feedback to clinicians and student users.

Existing research has documented the benefits of using mobile sensing technology for mental health research [7-10]. Among the pioneer projects, the MONARCA project used a variety of phone sensors, such as GPS (global positioning system) and an accelerometer, to detect the mental states of bipolar patients [11]. The StudentLife project used mobile sensing technology to monitor the daily behavior of college students and found that the tracked behaviors were associated with students' mental states, such as stress [12]. The CrossCheck project used data collected from phone sensors and ecological momentary assessment to build models to predict mental health indicators in schizophrenic patients based on phone sensor data [8]. In addition, mobile technologies have been shown to have advantages in delivering mental health therapies [13,14]. Systematic reviews have summarized the therapeutic effects of technology mediated mental health information systems [15]. For example, mental health apps on mobile phones could improve the accessibility to treatment and facilitate proactive seeking for professional help [13,16].

Building on previous research, this study represents the first effort toward applying a mobile sensing platform in enhancing mental health services at university counseling centers. The purpose of this study is to explore and gain initial insights into the usefulness and design of the mobile sensing platform using a qualitative approach. Specifically, the first purpose of this study is to understand the usefulness of a mobile sensing platform in improving counseling services and assisting students' self-management of their depression conditions.

Figure 1. An example of sensemaking behavioral targets [7]. GPS: global positioning system; SMS: short message service.



Moreover, a variety of behavioral targets could be extracted statistically using algorithms and machine learning [7]. However, domain expertise and human intelligence are crucial for constructing meaningful behavioral targets. Therefore, the second purpose of this study is to explore what types of behavioral targets and feedback are helpful from both the clinicians' and students' perspectives.

Methods

Participants

We recruited clinicians and student participants from the counseling center at Michigan State University (MSU). Recruitment occurred between February 15, 2017 and March 30, 2017. To recruit clinician participants, an invitation email with information about the study was sent to all 21 clinicians by the director of the counseling center and 9 clinicians responded to sign up for the study. To recruit student participants, a flyer was posted on the wall of the waiting room at the counseling center. To be an eligible participant in this study, students needed to meet the following criteria: (1) be 18 years old or older; (2) currently enrolled as a college student; (3) having been diagnosed with moderate, moderate-to-severe, or severe depression; and (4) are currently receiving college counseling services for their depression condition. Twelve eligible student participants signed up for the study. Each clinician and student participant received US $15 as compensation for participating the study. All participants had to sign a consent form in accordance with a study protocol approved by the MSU Institutional Review Board.

For the 9 clinicians (aged 31-55 years; mean 42 years, SD 5.83; 8 female), 3 were clinical psychologists, 3 were clinical counselors, 2 were educational psychologists, and 1 was a clinical social worker. For the 12 student participants (aged 19-22 years; mean 21, SD 1.22; 7 female), 5 were mildly depressed with PHQ-9 (PHQ-9 is a diagnostic assessment of major depression disorder) scores ranging from 9 to 14, and 7 were moderately depressed with PHQ-9 scores ranging from 15-19. Out of the 12 student participants, 10 were undergoing treatment (ie, antidepressants) with the student health center and all of them were receiving psychological counseling at the MSU counseling center or counseling service outside the MSU campus.

Procedures

We conducted semistructured interviews with all of the participants and the interviews lasted between 40 and 50 minutes. Interviews with clinicians were conducted in their offices; interviews with student participants were conducted at a location of their choice. The interviewees were informed that the background of the study was to develop and design a mobile sensing platform named iSee. The interviewees were informed that the iSee system leveraged the mobile sensing capacity of the mobile phone to collect raw sensor data related to physical properties of humans and environment. The iSee system then used data analytics to convert raw sensor data into meaningful behavioral targets, and then deliver behavioral targets information through a dashboard to clinicians and through a mobile app to students. For the interviews with the clinicians, the interviewer first described iSee and then displayed an example of a representation of sensed data from a student's phone that might be presented on a clinician's dashboard (Figure 2). For the interviews with the student participants, the interviewer first administrated PHQ-9 with paper and pencil to characterize the sample, then described iSee, and finally presented an example of output of sensed data on a mobile app (Figure 3). The interviewer explained to clinicians and students that the figures presented were only conceptual prototypes of iSee, and the purpose of the interview was to solicit their thoughts and opinion about the usefulness of iSee, what sensed behavioral targets they preferred, and how to present them via iSee.

After the introduction, three general lines of inquiry were pursued using semistructured interviews: (1) how the mobile sensing platform might be useful to counseling service (clinicians) and depression management (students), (2) what behavioral targets should be provided and in what format they should be provided to maximize the usefulness, and (3) identification of barriers to using the mobile sensing platform for clinicians and students (see interview protocols in Multimedia Appendix 1). All the interviews were audio recorded.

Figure 2. A conceptual prototype of iSee sensing platform and interface for clinicians.

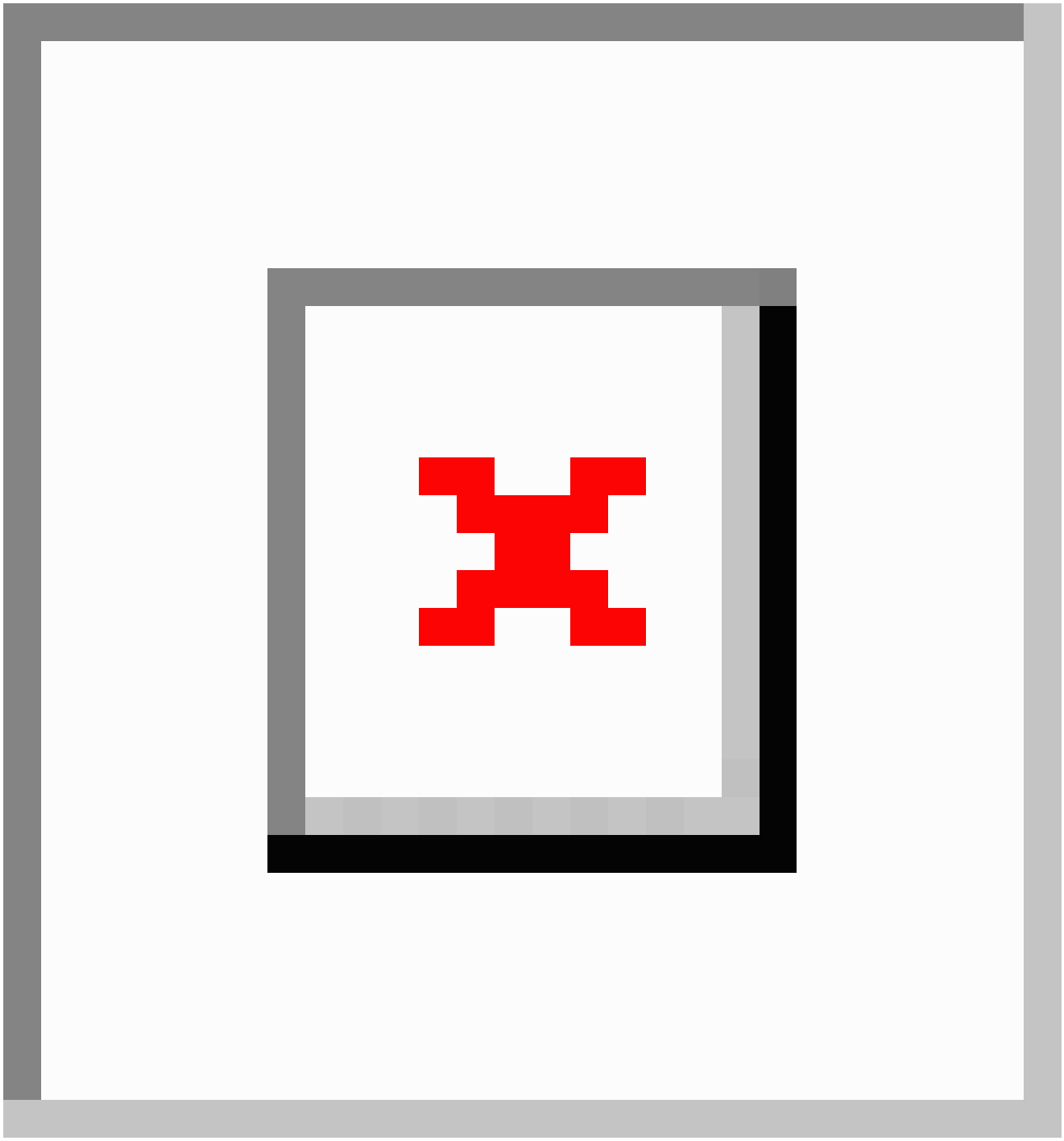
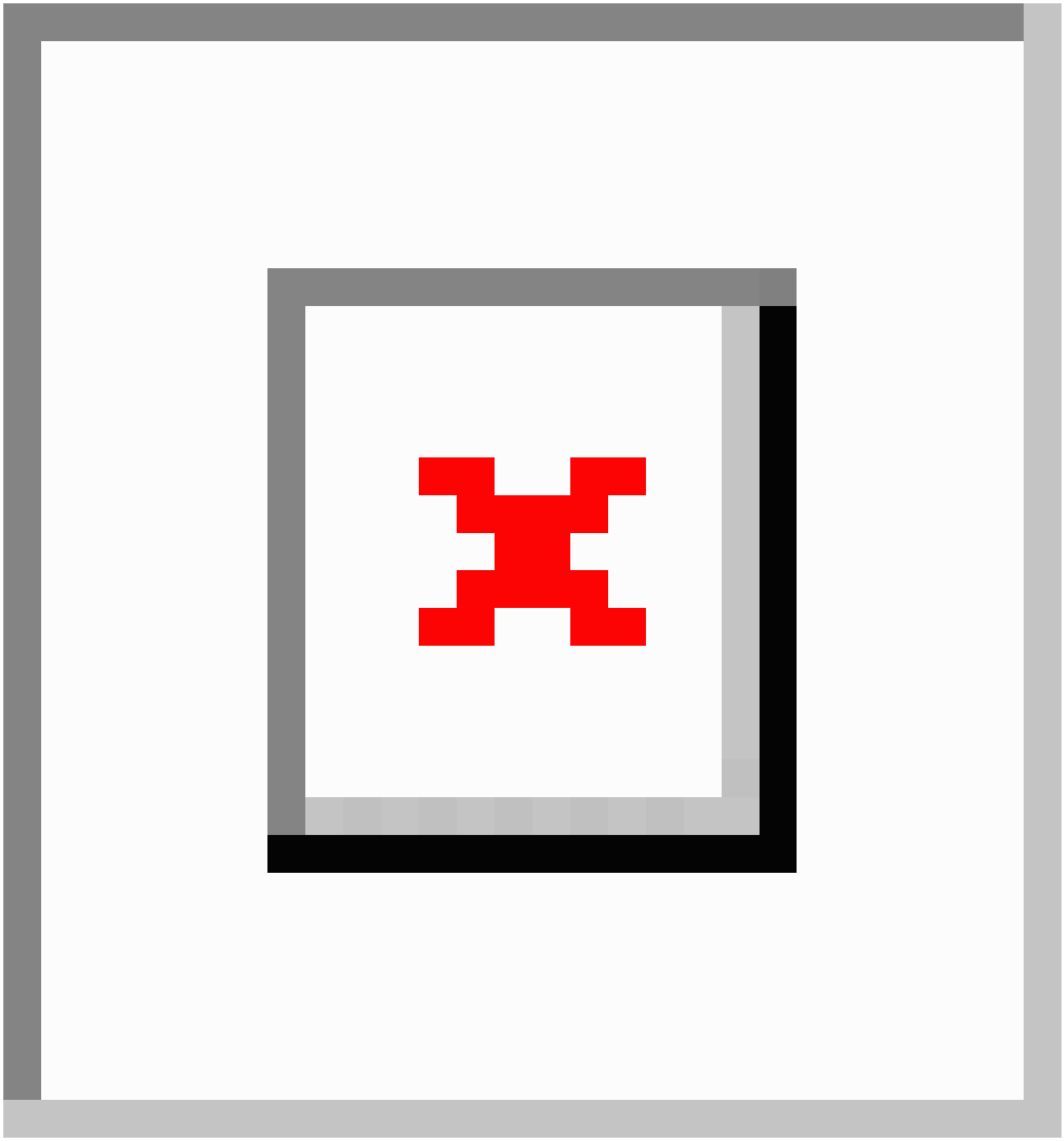


Figure 3. A conceptual prototype of iSee interface for students.



Data Analysis

All interviews were transcribed, and the interview data was analyzed using thematic analysis with an inductive approach [17]. Three researchers read the transcripts, familiarized themselves with the data, and independently generated an initial list of codes which represented the most basic elements in the raw data. In the next step, the same 3 researchers met frequently to discuss their initial codes, group initial codes into larger categories, and extracted the underlying concepts in the initial codes. In the subsequent analysis, they searched for themes, sorted the different categories and concepts into potential themes, and collated all the relevant coded data extracts within the identified themes using Nvivo 11. The last step of the analysis focused on reviewing all the collated extracts for each theme, considering whether they appeared to form a coherent pattern, and examining the validity of individual themes in relation to the entire dataset. This process involved iterative coding, memo writing, and constant comparison of the data to the emerging themes. The three researchers have background in health communication, which may affect their data interpretation and choices of themes.

Results

Clinician Perspective

We first report the findings from the clinician perspective with respect to the usefulness of iSee in improving college counseling

service for depression management, the preferred behavioral target data and feedback, and concerns about using iSee in counseling practice.

Usefulness of iSee in Counseling

Objective Feedback to Students with Depression

Seven clinicians believed that behavioral targets tracked via the mobile sensing platform would provide students with an objective check against their subjective beliefs about their behaviors.

> *...for people with depression, the subjective life experience is often different than the objective. You may think you're not eating enough, but you're eating too much; or you might think you've walked enough, but really you haven't, and instead, you've lain on bed for 14 hours.* [Clinician 1]

In particular, clinicians explained that depression could create a tendency to view things more negatively.

> *My clients always think negatively about themselves because that is what depression does. So, having the tracked data might help them do reality check. Like, you say have no friends and nobody likes you, but you have these many text messages exchanged with your friends.* [Clinician 4]

Objective Evidence to Clinicians

Seven clinicians said that they would assign homework (eg, increase physical activity) for students to complete between counseling sessions. Clinicians spoke about how the mobile behavioral-sensing platform could help them confirm students' compliance and adherence to the assigned homework.

> *I'm big on physical activity when treating depression...the data can help me to see if they are complying with the 45 minutes four days a week, then I can say confidently that we have done a serious intervention that would like parallel what would happen with a medication like an SSRI [selective serotonin reuptake inhibitor].* [Clinician 1]

Moreover, the mobile sensing platform can be useful to assist clinicians' decision-making about what homework to assign to students.

> *I'd like to see my clients' progress in completing the homework. Automatic tracking is helpful because I can go check and adjust the goals for them. Working towards small and achievable goals is very important.* [Clinician 9]

> *It's important to set an appropriate homework for them because if they cannot complete, they will see themselves fail, and that feeds their depression.* [Clinician 3]

Filling Communication Gaps

From the interviews, we learned that a counseling session typically ran for approximately 45-50 minutes. Due to the time limit, there may not be sufficient communication between clinicians and students. Five clinicians mentioned that, during one counseling session, they tended to focus on topics such as teaching mindfulness or dealing with a mental breakdown, leaving little time to communicate with students about their depressive symptoms such as sleep disruption and social avoidance. These clinicians said that the behavioral targets provided by the mobile sensing platform could complement clinician-student communication during the counseling session.

> *We have a progress note for each client, like asking about their sleep and eating when we start or end a counseling session. But sometimes we don't or forget to check progress because we are focusing on certain things like crises. But if it's a crisis, we really have to know the basic ones. The data can be a good reminder and it actually saves a lot of time for us.* [Clinician 3]

How to Provide Behavioral Targets

Data Provision to Students under Clinician Guidance

Three clinicians expressed a strong hope to have the mobile sensing platform developed and used under their guidance. This may help to maximize the benefits of showing data to students with depression while minimizing the potential risks.

> *If clients have never received counseling service before, and they might not have tolerance to reflect their own behaviors, telling them that their sleep quality is poor could be depressing. I would take a reserved stance in terms of how much and the kind of data given to clients, and when, depending on their conditions.* [Clinician 6]

Goal-Oriented Presentation of Behavioral Targets

Four clinicians suggested organizing behavioral targets around personalized goals. Clinicians would like to work with their clients to set up small and achievable goals (eg. go to bed at 11 pm and get up at 7 am) to manage sleep dysregulation in depression. Behavioral targets can be shown in terms of students' baseline and on-going progress towards their goals.

> *The best way to provide data to a client is to give them their baseline data, and then have them put in their goals, and then give them a readout at some regular interval of are you meeting your goal or exceed your goal.* [Clinician 1]

> *Clinicians can help to put an appropriate goal for a client. Instead of giving some scores out of context, it will be helpful to tell clients how they are doing compared to their baselines.* [Clinician 7]

How to Provide Feedback

Five clinicians pointed out that the value of the mobile sensing platform is that it can provide one's real-time behavior and constant behavioral targets, which serves as the basis of personalized feedback. A weekly counseling session takes up, at most, one hour, leaving 167 more hours per week where students with depression are on their own. Feedback based on students' tracked behavior could be delivered outside the counseling sessions by clinicians or the iSee platform. These

clinicians have mentioned that iSee could serve as an extension of counseling service beyond working hours.

Positive Reinforcement

The first type of feedback which emerged from our interview data is positive reinforcement in the form of esteem support messages, pointed out by 5 out of 9 clinicians. Esteem support messages defined as are words that acknowledge and validate one's self-worth and achievement [18]. When students are following clinicians' advice and making good progress in managing their depression conditions, positive reinforcement can boost their self-esteem and motivate them to continue.

> *I would like a portal for me to be able to go in and give them notes of encouragement. Like "Hey, John, saw your data today. You look great."* [Clinician 1]

Or if clinicians do not have time to review the behavioral target data, the system might automate message delivery based on the student's tracked behavior. For example, Clinician 8 suggested using the iSee platform to send the support message:

> *When clients are doing well, they may get an alert that says good job or congratulations. They will feel good about themselves.* [Clinician 8]

Reflection Reminder

Instead of showing students behavioral targets that might be demoralizing, 3 clinicians suggested that gentle reminders may be used to raise attention to a condition and encourage reflection.

> *When the tracked data do not look very good, you could use popup reminders that they can ask themselves questions. Like, "did you do things you need to do to take care of yourself today?" or "do you know that when your sleep is not well, you need to watch out because you might feel down this week?"* [Clinician 4]

These reminders may provide reassurance, activate students' reflection, or encourage preventive actions.

Challenging Negative Thoughts

Negative thoughts are a hallmark of depression. Clinician 9 pointed out that to review one's behavior trend could provide a context to challenge one's negative perception:

> *When students are depressed, they feel critical of themselves. Instead of saying, "Yay, it's great! I moved three times today." They are like, "I only moved three times today. This is the worst thing I'm feeling." So, at this point of time, we can provide alternative and more positive thinking.* [Clinician 9]

Barriers: Time and Liability

Time and liability are the two major barriers of using the mobile sensing platform at the counseling center. Two clinicians mentioned that they had very busy schedules, so they were not sure whether they would have time to review the data consistently. In addition, 3 clinicians were not entirely clear about their responsibilities to patients after having their behavioral targets information.

> *...it can be bad because if we missed something, or we can't reach them, then we are still responsible for the information that we have. It can go both ways, but it's just important to think about the liability piece.* [Clinician 3]

Student Perspective

In this section, we report the findings from the student perspective with respect to the usefulness of iSee in managing depression, the preferred behavioral targets and feedback, and concerns about using iSee in everyday life.

Usefulness: Support Continued Self-tracking

Nine out of the 12 student participants interviewed mentioned that they had used some kind of mobile tracking apps but unfortunately, it was difficult to continue the self-tracking practice. The discontinuance of self-tracking is mainly due to the amount of effort and time spent on inputting one's data.

> *I've tried tracking myself, but I stopped. The hardest part was logging the diet and exercise because there are so many variables that go into it.* [Student 4]
>
> *I used some trackers, but you had to input all of that data. I used that for a while but stopped because that's just like another thing to do.* [Student 7]

In particular, for students with depression, they may lack the motivation and energy to do self-tracking.

> *When I am going through a cycle of depression, I'm like, "oh, I'm so depressed now. Tracking is going to go by the wayside. It's just like not something on top of my list."* [Student 7]

Need for Data on Integrated Behavioral Targets

Seven out of the 12 student participants pointed out that they found data that could explain the correlations between various behavioral targets and their depression would be helpful to increase self-awareness.

> *I find nights that I noticeably don't sleep well, I have a worse mood when I wake up, so sleep tracking is important to me. I want to know how much I wake up versus how long I sleep, and how that affects me.* [Student 1]
>
> *I find running on treadmill and lifting weights is a way to relieve stress. It will be great if I can keep track of that and notice, hey, I feel happier today because I worked out.* [Student 10]
>
> *...it's good to kind of mentally be aware like, "wow, I ate this much food." I tend to eat more when I'm more depressed. It's more of a comfort food. Visualization will be helpful so that I can see really highs or lows.* [Student 11]

Avoid Negative Feedback, But Seek Positive Feedback

Four student participants tried to avoid seeing negative feedback because of the increased feeling of helplessness.

> *I feel like just get discouraged when you always are giving me the same results, or the same thing is*

happening. I want some type of improvement, but if isn't any, I don't want to know. [Student 12]

...isn't depressing to see my condition get worse? I would dismiss the data if I knew it would tell me something bad. [Student 8]

In contrast, 6 student participants expressed their love and need for positive supportive messages.

...some type of motivation or support and even if you had struggles, it wouldn't say "bad job," but "we know you had a tough day, but tomorrow will be better" or something. [Student 8]

...just a simple update of me is good enough. Like, "good job, keep up the good work" or "good improvement." [Student 4]

Barriers From a Student Perspective

Three out of the 12 student participants expressed their concerns about not being able to check on their behavioral targets on a regular basis. Student 7 stated:

When I'm depressed, I don't even want to take a look at the data or the problems, or what this could be. [Student 7]

Similarly, Student 8 did not feel confident to go through the data by himself.

I don't think I would check the tracked data all the time. It might be easier for me to go through the data with my doctor, so he can tell me what to do from there. [Student 8]

Table 1. Summary of themes from interviews with clinician and student participants.

Themes	Subthemes
Usefulness of the mobile sensing platform	
Clinician	• Objective feedback to students • Objective evidence to clinicians • Filling communication gaps
Student	• Support continued self-tracking
How to provide behavioral targets	
Clinician	• Provision under clinicians' guidance • Presentation oriented by personal goals
Student	• Preference of integrated behavioral targets
How to provide feedback	
Clinician	• Positive reinforcement • Reflection reminder • Challenge negative thoughts
Student	• Avoid negative but seek positive feedback

Table 1 provides a summary of major themes extracted from interviews with clinician and student participants.

During the interview, we prompted students with questions about privacy as a potential barrier of using iSee. Nine student participants did not feel privacy was an issue as long as their data were "anonymous." Eleven students were satisfied with the privacy issue when the researcher explained that iSee would deidentify personal data before transmitting to the cloud where the data would be stored. One student would like an option to choose what data to share:

...maybe some people want to make certain things private. Like I don't want people knowing how much I'm sleep. It would be a good idea to allow for that so people can make it function that way.

Discussion

Principal Findings and Implications

This study explored the types of behavioral targets, methods of delivering feedback, and user concerns with a mobile sensing platform, iSee, from both clinician and student perspectives. Our findings show that clinicians and students recognized the benefits of the mobile sensing platform in terms of providing objective behavioral data, filling clinician-student communication gaps, and easing continued self-tracking practice. Although clinicians and students shared thoughts on potential usefulness of iSee, they differed on preferences for the types of behavioral targets and types of feedback.

Individual Versus Integrated Behavioral Targets

Clinician and student participants expressed different views on how to provide behavioral targets. Clinicians emphasized presenting individual behavioral targets in relation to corresponding behavioral goals, whereas students would like to see integrated behavioral targets in relation to their depression

conditions. Clinicians saw value in setting up appropriate behavioral goals that are small and achievable, such as exercising 30 minutes three days a week or sleeping for 8 hours every night. They were interested in using visualizations of sensed behavioral targets to observe how much progress students have made from their baseline behaviors to the specific goals over time. In comparison, student participants wanted to understand the correlations between clusters of behaviors and their distress. For example, a student participant may be more interested in knowing how waking up during the night and lack of deep sleep are related to a depressed feeling next day than knowing that she slept only 4 hours yesterday. This requires a juxtaposition of several behavioral targets (eg, wake up, deep sleep, sleep length) and their associations with levels of depression.

The different viewpoints between clinicians and students may be due to their different objectives, which may require different design approaches. Clinicians frequently make behavioral recommendations, and much of treatment involves encouraging patients to make those changes. Thus, tools that support such monitoring may support clinicians in being more effective in their roles. According to existing literature, students are interested in self-experimentation and learn self-management skills from personal experience [19,20,21]. Therefore, for students who may not be fully aware of the associations between their behaviors and depression, it is critical to provide integrated data visualizations to facilitate sense making of behavioral targets. For students who have the awareness of how their behaviors affect depression, they might be more motivated to set up specific behavioral goals and review relevant behavioral targets for achieving the goals. While the goals of clinicians and student participants may require different design approaches, they are compatible.

Positive Versus Negative Feedback

Both clinicians and students embraced positive feedback when a student was making good progress. Messages that confirmed one's achievement, validated one's self-worth, and encouraged continued efforts were welcomed. However, clinicians and students differed to some extent in terms of whether or not providing feedback when students were not making positive progress. Students generally declined to review the behavioral targets when it did not show any positive change because the situation may make them feel depressed. This finding was consistent with existing literature on self-trackers' experiences. For example, one study found that self-trackers experienced frustration and anxiety when they were aware of negative tracked data [20].

Clinicians presented more diverse opinions about this issue. Some clinicians suggested using gentle reminders or reflection questions to raise students' attention to the issue reflected in negative data. Some clinicians suggested only showing negative data to students during counseling sessions so that clinicians could discuss the data with students. As suggested in the self-tracking literature, designers could build tools that customize the individual user experience [22]. Applied to the current study, the platform could solicit students' preferences with regard to sharing and discussing negative data with their clinicians, receiving gentle reminders without presenting the actual behavioral data, or dismissing any negative information. Future research is encouraged to ask perspective-taking questions so that clinicians and students could stand in each other's shoes and see whether a more shared perspective could emerge.

Concerns About Reviewing Data

While iSee attempts to reduce the burden on students of entering data, both students and clinicians nevertheless saw reviewing the data as a potential burden. Clinicians expressed concerns about not having enough time to review students' tracked data, while students were worried that they might not be able to check on their data when they were too depressed. Some students preferred that their clinicians take the primary role of checking on their data. These concerns reflect a barrier of effectively using iSee in the clinical setting: some uncertainty in the capacity to pay attention to behavioral targets.

A couple of steps could be taken to resolve this potential barrier. For clinicians, designers are encouraged to decrease the overhead of reviewing data and using iSee. More research should be conducted to establish a stronger confidence in behavior targets that are clinically meaningful so that clinicians can make quick relevance of the data [7]. While clinicians are protective of their time, they would spare time on the platform if they believe it is beneficial to counseling [23]. In addition, sending encouragement and motivational messages may not be a good use of clinicians' expertise [23]. Therefore, iSee could automate the process of sending supportive messages based on users' tracked data. For students, designers might consider using push methods, such as sending data visualizations and feedback via text messages [24], instead of expecting students to open the app and review it. Moreover, tailoring data review to students' needs (eg, what to review, how frequently to review) could give students a sense of ownership and control [23] that enhances students' engagement with the system.

Limitations and Future Research

Our study has several limitations. First, the study sample is small, and all participants were recruited from only one university counseling center. Caution should be taken when generalizing the interview results to clinicians and students from other university counseling centers and college students with depression. Nevertheless, the study site MSU is a large public university in the US; and, therefore, should share important characteristics with counseling centers, students, and clinicians from other public universities.

Another limitation is that our sample did not include any students with severe depression. More severely depressed students may have potentially different opinions regarding the use of iSee.

The third limitation of the study is the lack of field deployment. Given the exploratory nature of the study, we only presented the conceptual prototype of iSee to interviewees to solicit their thoughts and opinions. User experience may vary when they use the actual platform. For example, the quality of collected personal data could depend on the system design and implementation in the field deployment. The benefits of iSee

and useful types of behavioral target data and feedback will have to be validated once the iSee is developed and deployed. The effectiveness of design features can only be evaluated in a randomized controlled trial of the iSee system.

This study has explored potential feedback to deliver to students with depression (eg, positive feedback, reflection reminders). The next stage of research could examine how the sensor data and behavioral targets could inform the design and delivery of micro-interventions. For example, how frequently and when should an encouraging message be sent to the user? This will require extensive user testing of different design ideas and prototypes. Future work is encouraged to connect sensor data, behavioral targets, feedback, and interventions.

Finally, we also note that this study focused only on behavioral targets where there is sufficient evidence that they can be accurately sensed using mobile phone sensors. While it is possible that these findings may extend to self-reported data, we would caution against extending this to behavioral targets not explored in this study.

Conclusion

By conducting interviews with clinicians and student participants, we have explored the issues surrounding benefits of iSee, and useful types of behavioral target data and feedback. We have gained some initial insights such that the behavioral data generated by iSee could complement students' activities and behaviors self-reported during counseling sessions, fill in clinician-student communication gaps, and extend therapy beyond the clinical settings by delivering appropriate feedback. With respect to preferred types of behavioral targets, we have learned that clinicians may focus on individual behavior targets with a set goal, whereas students may prefer integrated behavioral targets that assist their understanding of the relationship between their behaviors and depression. In addition, clinicians may have diverse opinions about presenting negative data to patients, whereas students try to avoid negative feedback. This qualitative work represents the first effort to understand the benefits and user needs of a mobile sensing platform such as iSee in university counseling service.

Acknowledgments

This research is funded by NSF award number 1632051 and is supported by Microsoft Research. We thank Ying Cheng for her help with coding interview scripts to establish intercoder reliability.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview protocols used for clinician and student participants.

[PDF File (Adobe PDF File), 398KB - jmir_v20i7e10139_app1.pdf]

References

1. American College Health Association. American College Health Association-National College Health Assessment Spring 2007 Reference Group Data Report (Abridged). Journal of American College Health 2008 Mar;56(5):469-480. [doi: 10.3200/jach.56.5.469-480]
2. Buchanan JL. Prevention of depression in the college student population: a review of the literature. Arch Psychiatr Nurs 2012 Feb;26(1):21-42. [doi: 10.1016/j.apnu.2011.03.003] [Medline: 22284078]
3. Reetz D, Bershad C, LeViness P, Whitlock M. www.aucccd.org. 2016. The association for university and college counseling center directors annual survey URL: https://www.aucccd.org/assets/documents/aucccd%202016%20monograph%20-%20public.pdf [accessed 2018-07-05] [WebCite Cache ID 70giMvNZs]
4. Miller-Matero LR, Clark KB, Brescacin C, Dubaybo H, Willens DE. Depression and literacy are important factors for missed appointments. Psychol Health Med 2016 Sep;21(6):686-695. [doi: 10.1080/13548506.2015.1120329] [Medline: 26695719]
5. Haerizadeh M, Moise N, Chang B, Edmondson D, Kronish I. Depression and doctor-patient communication in the emergency department. Gen Hosp Psychiatry 2016;42:49-53 [FREE Full text] [doi: 10.1016/j.genhosppsych.2016.06.004] [Medline: 27638972]
6. Schiepek G, Aichhorn W, Gruber M, Strunk G, Bachler E, Aas B. Real-Time Monitoring of Psychotherapeutic Processes: Concept and Compliance. Front Psychol 2016;7:604 [FREE Full text] [doi: 10.3389/fpsyg.2016.00604] [Medline: 27199837]
7. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017 May 08;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
8. Wang R, Aung M, Abdullah S, Brian R, Campbell AT, Choudhury T, et al. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. 2016 Presented at: Proceedings of the ACM International

Joint Conference on Pervasive and Ubiquitous Computing - UbiComp 16; 2016; Heidelberg, Germany p. 886-897. [doi: 10.1145/2971648.2971740]
9. Saeb S, Zhang M, Kwasny M, Karr C, Kording K, Mohr D. The Relationship between Clinical, Momentary, and Sensor-based Assessment of Depression. In: Int Conf Pervasive Comput Technol Healthc. 2015 Aug Presented at: Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2015 9th International Conference on (pp.). IEEE; 2015 May 20; Istanbul, Turkey p. 229-232 URL: http://europepmc.org/abstract/MED/26640739 [doi: 10.4108/icst.pervasivehealth.2015.259034]
10. Saeb S, Zhang M, Karr C, Schueller S, Corden M, Kording K, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273]
11. Gravenhorst F, Muaremi A, Bardram J, Grünerbl A, Mayora O, Wurzer G, et al. Mobile phones as medical devices in mental disorder treatment: an overview. Pers Ubiquit Comput 2014 Sep 21;19(2):335-353. [doi: 10.1007/s00779-014-0829-5]
12. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014 Presented at: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp 14 Adjunct; 2014; Seattle, Washington p. a. [doi: 10.1145/2632048.2632054]
13. Donker T, Straten AV, Marks I, Cuijpers P. A Brief Web-Based Screening Questionnaire for Common Mental Disorders: Development and Validation. J Med Internet Res 2009 Jul 24;11(3):e19. [doi: 10.2196/jmir.1134]
14. Riper H, Andersson G, Christensen H, Cuijpers P, Lange A, Eysenbach G. Theme issue on e-mental health: a growing field in internet research. J Med Internet Res 2010 Dec;12(5):e74 [FREE Full text] [doi: 10.2196/jmir.1713] [Medline: 21169177]
15. Wahle F, Bollhalder L, Kowatsch T, Fleisch E. Toward the Design of Evidence-Based Mental Health Information Systems for People With Depression: A Systematic Literature Review and Meta-Analysis. J Med Internet Res 2017 May 31;19(5):e191 [FREE Full text] [doi: 10.2196/jmir.7381] [Medline: 28566267]
16. Kim J, Lim S, Min YH, Shin Y, Lee B, Sohn G, et al. Depression Screening Using Daily Mental-Health Ratings from a Smartphone Application for Breast Cancer Patients. J Med Internet Res 2016 Aug 04;18(8):e216 [FREE Full text] [doi: 10.2196/jmir.5598] [Medline: 27492880]
17. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
18. Cohen S, Hoberman HM. Positive Events and Social Supports as Buffers of Life Change Stress1. J Appl Social Pyschol 1983 Apr;13(2):99-125. [doi: 10.1111/j.1559-1816.1983.tb02325.x]
19. Karkar R, Zia J, Vilardaga R, Mishra SR, Fogarty J, Munson SA, et al. A framework for self-experimentation in personalized health. J Am Med Inform Assoc 2016 May;23(3):440-448. [doi: 10.1093/jamia/ocv150] [Medline: 26644399]
20. Choe EK, Lee N, Lee B, Pratt W, Kientz JA. Understanding quantified-selfers' practices in collecting and exploring personal data. 2014 Presented at: the SIGCHI Conference on Human Factors in Computing Systems - CHI2014; April 26 - May 01, 2014; Toronto, Canada.
21. Ravichandran R, Sien S, Patel SN, Kientz JA, Pina LR. Making sense of sleep sensors: How sleep sensing technologies support and undermine sleep health. 2017 Presented at: The CHI Conference on Human Factors in Computing Systems - CHI 17; May 06- May 11, 2017; Denver, Colorado. [doi: 10.1145/3025453.3025557]
22. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
23. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. 2012 Presented at: SIGCHI conference on Human Factors in Computing System; May 05 - May 10; Austin, Texas p. 1421-1430.
24. Bardram JE, Frost M, Szanto K, Vinberg M, Kessing LV. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. 2013 Presented at: the SIGCHI conference on human factors in computing system; April 27 - May 02; Paris, France p. 2627-2636. [doi: 10.1145/2470654.2481364]

Abbreviations

GPS: global positioning system
MSU: Michigan State University

Edited by R Calvo, J Torous, G Wadley; submitted 15.02.18; peer-reviewed by I Montagni, M Rabbi, M Birk, V Klaas; comments to author 08.03.18; revised version received 01.05.18; accepted 10.06.18; published 17.07.18.

Please cite as:
Meng J, Hussain SA, Mohr DC, Czerwinski M, Zhang M
Exploring User Needs for a Mobile Behavioral-Sensing Technology for Depression Management: Qualitative Study
J Med Internet Res 2018;20(7):e10139
URL: http://www.jmir.org/2018/7/e10139/
doi:10.2196/10139
PMID:30021710

Original Paper

Using Mobile Phone Sensor Technology for Mental Health Research: Integrated Analysis to Identify Hidden Challenges and Potential Solutions

Tjeerd W Boonstra[1], MPsych, PhD; Jennifer Nicholas[1], BSc, BA, PhD; Quincy JJ Wong[1,2], MPsych, PhD; Frances Shaw[1], MA, PhD; Samuel Townsend[1], BE, MBiomEng; Helen Christensen[1], MPsych, PhD

[1]Black Dog Institute, University of New South Wales, Sydney, Australia

[2]Western Sydney University, Sydney, Australia

Corresponding Author:
Tjeerd W Boonstra, MPsych, PhD
Black Dog Institute
University of New South Wales
Hospital Road
Sydney,
Australia
Phone: 61 2 9382 9285
Email: t.boonstra@unsw.edu.au

Abstract

Background: Mobile phone sensor technology has great potential in providing behavioral markers of mental health. However, this promise has not yet been brought to fruition.

Objective: The objective of our study was to examine challenges involved in developing an app to extract behavioral markers of mental health from passive sensor data.

Methods: Both technical challenges and acceptability of passive data collection for mental health research were assessed based on literature review and results obtained from a feasibility study. Socialise, a mobile phone app developed at the Black Dog Institute, was used to collect sensor data (Bluetooth, location, and battery status) and investigate views and experiences of a group of people with lived experience of mental health challenges (N=32).

Results: On average, sensor data were obtained for 55% (Android) and 45% (iOS) of scheduled scans. Battery life was reduced from 21.3 hours to 18.8 hours when scanning every 5 minutes with a reduction of 2.5 hours or 12%. Despite this relatively small reduction, most participants reported that the app had a noticeable effect on their battery life. In addition to battery life, the purpose of data collection, trust in the organization that collects data, and perceived impact on privacy were identified as main factors for acceptability.

Conclusions: Based on the findings of the feasibility study and literature review, we recommend a commitment to open science and transparent reporting and stronger partnerships and communication with users. Sensing technology has the potential to greatly enhance the delivery and impact of mental health care. Realizing this requires all aspects of mobile phone sensor technology to be rigorously assessed.

(J Med Internet Res 2018;20(7):e10131) doi:10.2196/10131

KEYWORDS

passive sensing; mental health; ubiquitous computing; ethics; depression; mobile health; smartphone; wearable sensors

Introduction

Background

Mobile phone sensor technology has great potential in mental health research, providing the capability to collect objective data on behavioral indicators independent of user input [1-3]. With the plethora of sensors built into mobile phones, passive collection of a wide range of behavioral data are now possible using the device most people carry in their pockets [4]. Passive data collection operates in the background (requires no input from users) and allows measurement of variables longitudinally with detailed moment-to-moment information and collection

of temporal information on dynamic variables, such as users' feelings and activity levels. Given that these digital records reflect the lived experiences of people in their natural environments, this technology may enable the development of precise and temporally dynamic behavioral phenotypes and markers to diagnose and treat mental illnesses [5].

An ever-growing number of mobile phone apps have been developed to passively collect sensor data for mental health purposes, for example, Purple Robot is a mobile phone sensor data acquisition platform developed at Northwestern University that is available on Android devices. The platform gives access to a range of sensors including device information, battery level, Bluetooth, Wi-Fi, global positioning system (GPS), accelerometer, and communication logs [6] and has been used in research studies on mental health in the general community [7,8]. The Beiwe Research Platform was developed at Harvard University to collect research-quality data from mobile phone sensors, including use patterns, on both Android and iOS platforms in primarily clinical samples. The app collects a range of sensor data including that obtained from GPS, accelerometer, communication logs, Wi-Fi and Bluetooth (Android only), and battery use [9]. Another notable example is the large-scale Copenhagen Networks Study, a research project studying social networks in 1000 university students, which provided Android mobile phones to collect Bluetooth, Wi-Fi, communication logs, and GPS sensor data [10]. These different software and methodological approaches have resulted in different behavioral indicators being targeted, different features extracted, and different statistical approaches used to link behavioral data to mental health.

Behavioral Markers of Mental Health

Depression is associated with a number of behavioral changes, of which sociability and activity are the most studied using mobile phone sensor data. Social connectedness is a key factor in mental health and well-being. Social isolation, perceptions of loneliness, lower perceived social support, and fewer close relationships have all been linked to depression [11,12]. Mental health is also affected by the location of individuals within their social network and the number and strength of their social connections [13]. Traditionally, social networks have been investigated using self-reported data, the reliability of which has been questioned [14]. Sensor-enabled mobile phones allow for the collection of passive data to map social networks of proximity using sensor data including that obtained from Bluetooth. Eagle et al [15] were able to differentiate friends from nonfriends accurately using temporal and spatial patterns of Bluetooth data. As far as we know, mobile phones that collect Bluetooth sensor data have not yet been used in mental health studies. However, Pachucki et al [16] have used wearable proximity sensors to map a social network in high school children, showing that adolescent girls with more depressive symptoms have smaller social networks.

Depression is also associated with decreased activity and motivation and increased sedentary behavior [17]. Cross-sectional data indicates that people with depression are less likely to be active than people without depression [18]. Furthermore, longitudinal studies have shown that baseline depression is associated with increased sedentary behavior over time [18] and that low physical activity at baseline is associated with increased depression [19]. Again, mobile phone sensors, particularly GPS, are well placed to monitor an individual's location, physical activity, and movement. Initial research in a small sample (N=18) has indicated potential features of GPS data, such as a lower diversity of visited places (location variance), more time spent in fewer locations, and a weaker 24-hour, or circadian, rhythm in location changes, that are associated with more severe depression symptoms [7].

Challenges of Mobile Phone Sensor Technology

Despite the potential of mobile phone sensor technology in mental health research, this promise has not yet been brought to fruition. The use of mobile phone sensor technology for mental health research poses several key challenges, both technical and issues specific to mental health apps. A primary technical challenge is the reliable collection of sensor data across mobile platforms and devices, for example, location data may be missing due to sensor failure to obtain GPS coordinates [20,21], participants not charging or turning off their phones, or unavailability of any network connections for a long period of time, hampering data transfer to servers [7,10]. The mode of data collection also influences data completeness, which can differ between operating systems. Passive collection of sensor data are easier to support on Android than iOS; about twice as many apps are available for Android than for iOS [22]. This likely reflects greater restrictions that iOS places on accessing system data and background activity, making personal sensing using iOS devices challenging.

Another technical issue is battery life. Frequent sampling of sensor data can consume a significant proportion of a mobile phone's battery [23]. Ultimately, if an app collecting sensor data are too resource-intensive, users' motivation to continue using it decreases [24], which may lead to the app being uninstalled, ceasing the flow of data to researchers. Optimizing passive data collection to obtain the most detailed information possible should therefore be balanced with expectations of users regarding battery consumption. This is a significant practical challenge faced by mobile sensing apps.

In addition, there are specific challenges for using mobile phone sensor technology for mental health purposes, such as the engagement and retention of users [25]. Increasingly, a user-centered design approach is considered an integral part of any mental health app development [26-29]. Individuals with the target disorder can provide important information about the direction and focus of the app as well as how they engage with an app given their symptom profile. For example, focus groups of individuals with Posttraumatic Stress Disorder (PTSD) indicated that PTSD Coach was particularly useful for managing acute PTSD symptoms and helping with sleep [30]. Clinicians, on the other hand, can provide input into the design and functionality of an app from a therapeutic perspective. For example, clinicians indicated that an app for individuals with bipolar disorder to self-manage their symptoms should focus on medication adherence, maintaining a stable sleep pattern, and staying physically and socially active [31]. Codesign of mental health apps with end users and other stakeholders

increases the likelihood that the app will be perceived as attractive, usable, and helpful by the target population [24]. Although design and usability issues are often discussed for apps that require active user engagement, it is also important for passive data collection apps to increase user engagement and retention because this will ensure lower rates of missing data and dropouts. Furthermore, many apps have an ecological momentary assessment (EMA) component to complement passive sensor data collection.

User perceptions of an app's confidential handling and use of data, as well as privacy and anonymity, are additional challenges of passive data collection [9,32,33]. Mental health data are highly sensitive because of the potential negative implications of unwanted disclosure [34]; therefore, uncertainty about whether a service is confidential can be a barrier to care [35]. Indeed, data privacy and confidentiality are major concerns for the users of mental health apps [36,37], but no consensus has yet been reached on ethical considerations that need to be addressed for the collection of passive sensor data. Moreover, user perceptions of security and privacy may differ; for example, Android and iOS users differ in characteristics such as age and gender [38] and also in their awareness about security and privacy risks of apps [39]. Deidentification may be used to the protect privacy of individuals [40] but may also remove information that is important to maintain the usefulness of data, depending on context and purpose for use [41]. Systems making use of predictive analysis techniques not only collect data but also *create* information about personal mental health status, for example, through identification of markers for risk [42]. Therefore, social impact needs to be considered beyond individual privacy concerns.

Outline

In this study, we examined challenges of using mobile phone sensor technology for mental health research by analyzing results of a feasibility study that was conducted to test an app collecting passive sensor data. We analyzed the amount of sensor data that was collected, assessed the ability to quantify behavioral markers from Bluetooth and GPS data collected in a real-world setting, quantified battery consumption of the app, and examined user feedback on usability. No mental health questionnaires were administered as part of the feasibility study, although demographic and diagnostic data were available from the volunteer research register from which participants were drawn. We also investigated views of participants about acceptability of passive data collection for mental health research. The purpose of collecting this information was to build greater understanding of how social norms and perceptions around technology and data collection impact the feasibility, ethics, and acceptability of these technologies. We related results from our feasibility study to existing literature in these areas to identify common challenges of using mobile phone sensor technology in mental health research. We also drew some distinctions between available apps and made brief recommendations for the field going forward.

Methods

Mobile Phone App

Socialise, a mobile phone app developed at the Black Dog Institute, was used to assess the feasibility and challenges of passive data collection in a group of volunteers. We developed Socialise as a native app in Java for Android and Objective-C for iOS to collect passive data (Bluetooth and GPS) and EMA. Building on the results of a previous validation and feasibility study [43,44], we implemented several changes to improve scanning rates on iOS and here we tested Socialise version v0.2. We used silent push notifications to trigger Bluetooth and GPS scans and to upload data to the server. Silent push notifications, along with the "content-available" background update parameter, were used to deliver a payload containing an operation code corresponding to either a Bluetooth or GPS scan or one of a number of data uploads. The allowable background time for processing a push notification is sufficient to perform these scans and record data, and we hence used silent push notification to overcome some of the limitations imposed by iOS on apps running in the background. In addition, we used the significant-change location service to improve data collection rates. Unlike Android devices, no mechanism exists on iOS to allow the app to relaunch when a device restarts. By subscribing to the significant-change location service, the app is notified when the device restarts and triggers a local notification reminding participants to resume data collection.

Participants and Procedure

This study was approved by the University of New South Wales Human Research Ethics Committee (HC17203). Participants were recruited through advertisements disseminated through the Black Dog Institute volunteer research register. Individuals sign up on this register to volunteer for research. As part of the sign-up process, individuals provide demographics and diagnostic information (ie, mental disorders they have experienced in their lifetimes). To be able to participate in this study, individuals had to be 18 years or older, reside in Australia, speak English, and have a mobile phone running Android version 4.4 or newer or running iOS8 or newer. Interested individuals received a link to the study website where they could read participant information and provide consent. Of the 32 participants who provided consent to participate in the study, 31 also agreed to have their data made available on a public repository. Once they gave consent, participants received a link to install the Socialise app and a unique participant code. When participants opened the app, they were asked to give permission for the app to receive push notifications and collect location and Bluetooth data. Participants then had to fill in the unique participant code. Once the app opened, participants were asked to complete an entry survey, which included questions about the age of their mobile phone, the amount of time spent on their phone each day, and evaluation of their satisfaction with the onboarding process.

Participants were instructed to use the Socialise app for 4 weeks. Bluetooth and GPS data were collected during scans that were conducted at intervals of 8, 5, 4, or 3 minutes (equivalent to 7.5, 12, 15, and 20 scans per hour, respectively). Each scanning rate

was tested for 1 week, and participants were instructed to use their phones normally for the duration of the study.

Data Collection

We used the *BluetoothManager* private API on iOS devices to collect Bluetooth data, because the public *CoreBluetooth* API contains only functions for interacting with low-energy devices. It is currently not feasible to use Bluetooth Low Energy to map social networks in iOS [45]. To collect GPS data, the *CoreLocation* framework was utilized on iOS. The Android implementation leveraged the built-in Bluetooth APIs and *LocationManager* to collect Bluetooth and GPS data. Data acquisition settings were identical on iOS and Android, and both were set to collect Bluetooth, GPS, and battery data every 3, 4, 5, and 8 minutes.

Because the Bluetooth media access control address of a device is potentially personally identifiable information, these data were cryptographically hashed on the handset to ensure the privacy of participants. Hashing generates a consistent "signature" for each data item that cannot be reversed to reveal the original data value. To record only other mobile phones, detected devices were filtered according to the Bluetooth Core Specification. This involved removing any devices not matching the Class of Device 0×200 during the Bluetooth scan.

Participants were asked to complete a short questionnaire at the end of each week to document any problems that they encountered using the app. It included questions about whether they had changed phone settings (eg, turned off GPS or mobile data or turned on airplane mode), whether they used Bluetooth on their phone, and whether they thought the Socialise app impacted battery life. These findings were evaluated using a 7-point Likert scale. In addition, a set of questions about the acceptability of sensor data collection and some contextual information about that acceptability was collected at the end of the study.

Data Analysis

Data completeness was assessed by comparing the number of Bluetooth and GPS scans that were scheduled for the duration of the study (9156 samples per participant) with the number of data samples that were uploaded by the app; that is, we scheduled scans every 3, 4, 5, and 8 minutes, each for a week (4 weeks), which comes to 20 24 7 + 15 24 7 + 12 24 7 + 7.5 24 7=9156 total scans.

Most research using mobile phone Bluetooth to track social interactions has been performed in closed social networks [10,15,43,46]. In contrast, in this study, sensor data were collected from participants living in Australia who were unlikely to have social connections with each other. We therefore followed procedures described by Do et al [47] for analyzing Bluetooth data in a real-world setting. Instead of using Bluetooth to assess social connection between participants, Bluetooth was used to make a coarse estimate of human density around the user, which provides a rough proxy for social context. We first distinguished between known and unknown devices. Known devices were defined as devices that had been observed on at least 3 different days during the duration of the study. We then computed the average number of known and unknown devices that were detected at each hour of the day to obtain a social context profile for each participant.

We followed procedures outlined in Saeb et al [7] for analyzing GPS data. To identify location clusters, we first determined whether each GPS location data sample came from a stationary or a transition state. We calculated the time derivate to estimate movement speed for each sample and used a threshold of 1 km/h to define the boundary between the two states. We then used *K*-mean clustering to partition data samples in the stationary state into K clusters such that overall distances of data points to centers of their clusters were minimized. We increased the number of estimated clusters from 1 until the distance of the farthest point in each cluster to its cluster center fell below 500 m. We also estimated circadian movement, a feature that strongly correlated with self-reported depressive symptom severity [7]. Circadian movement measures to what extent participants' sequence of locations follows a 24-hour rhythm. To calculate circadian movement, we used least squares spectral analysis [48] to obtain the spectrum of GPS location data and estimate the amount of energy that fell with the 24-hour frequency bin. Circadian movement was then defined as the logarithm of the sum of energy for longitude and latitude [7].

The battery consumption of the Socialise app was estimated by varying the scanning rate each week. Varying scan rates enabled us to differentiate the battery consumption of the Socialise app from that of other apps running on participants' mobile phones. We estimated the battery consumption of the Socialise app using linear regression, assuming that battery consumption scaled linearly with the number of scans performed per hour. To estimate battery consumption, we first extracted data samples when the battery was discharging and then computed the change in battery charge between scans. We then estimated the length of time for the battery to be exhausted separately for each scanning rate and device. We used a robust fitting algorithm, that is, reweighted least squares with the bisquare weighting function [49], to estimate the average battery consumption across devices and how it changed with scanning rate.

All analyses were performed using Matlab version R2018a (The MathWorks Inc, Natick, MA, USA).

To evaluate user perceptions of battery consumption of the app, we compared responses on perceived impact on battery life across the 4 weeks of the study to assess whether perceived impact was affected by the actual scanning rate. To examine views of participants about the acceptability of passive data collection for mental health research, we compared their responses for different data types and contexts using a one-way repeated-measures analysis of variance (ANOVA). Statistical analyses were performed using JASP version 0.8.3.1 (University of Amsterdam, Netherlands). We also collected open responses to these questions, allowing for qualitative analysis. However, owing to the small number of responses, coding to saturation was not possible and we conducted a thematic analysis instead, dividing responses into categories to determine their approximate range.

Results

Participant Characteristics

Overall, 53 people expressed interest in participating in the study. Of these, 41 completed registration and gave informed consent. Of the 41, 1 participant was not eligible because the person did not live in Australia, 1 participant withdrew, 2 participants were unable to install the app on their mobile phones, and 5 participants did not respond to the follow-up email. The remaining 32 participants successfully installed the app on their mobile phones. The age of participants was broadly distributed with the majority aged from 55 to 64 years (see Table 1). Most were female (23/30, 77%) and reported that they had been diagnosed with a mental disorder (23/32, 72%); depression and anxiety disorders were most commonly reported (Table 1).

Participants reported using their mobile phones regularly, and most devices were less than a year old (15/30, 50%).

Data Completeness

Over the course of the study, 1 participant withdrew and another stopped participating. We therefore obtained sensor data from 28 of the 41 who consented to participate with a retention rate of 68%. Survey data were collected from 23 participants (participants who provided at least one response on the short questionnaire at the end of each week) and 13 participants completed the exit survey, as seen in Figure 1. Over the 4 weeks, a total of 9156 data points was scheduled for each participant. On average, 55 (19%) of scheduled samples were collected on Android and 45 (20%) on iOS, as seen in Figure 2. The figure shows the percentage of the number of scheduled samples that were collected on the devices used in the study. The x-axis lists the mobile phone model that each participant used.

Table 1. Participant demographics.

Characteristics	n (%)
Sex (n=30)	
Male	7 (23)
Female	23 (77)
Age in years (n=30)	
18-24	5 (17)
25-34	6 (20)
35-44	5 (17)
45-54	4 (13)
55-64	7 (23)
65+	3 (10)
Mental disorder diagnosis (n=32)	23 (72)
Depression	22 (69)
Bipolar disorder	9 (28)
Anxiety disorder	17 (53)
Schizophrenia	0 (0)
Personality disorder	2 (6)
Substance use disorder	5 (16)
Eating disorder	7 (22)
Autism spectrum disorder	1 (3)
Posttraumatic Stress Disorder	2 (6)
Attention deficit hyperactivity disorder	1 (3)
Daily phone usage (n=30)	
Less than 30 min	2 (7)
30 min-1 h	7 (23)
1-2 h	4 (13)
2-3 h	6 (20)
More than 3 h	11 (37)

Figure 1. Flowchart of the number of participants entering the study.

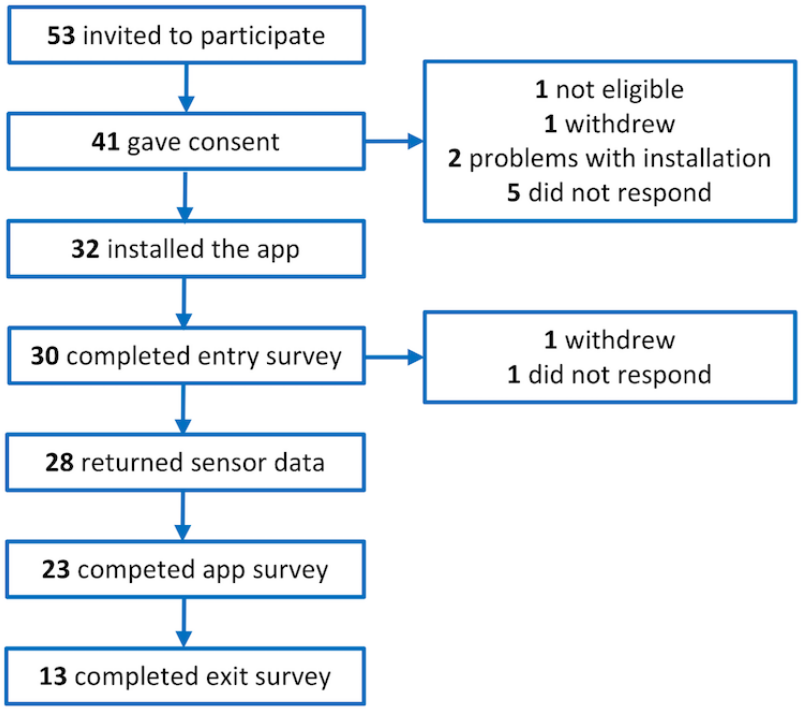
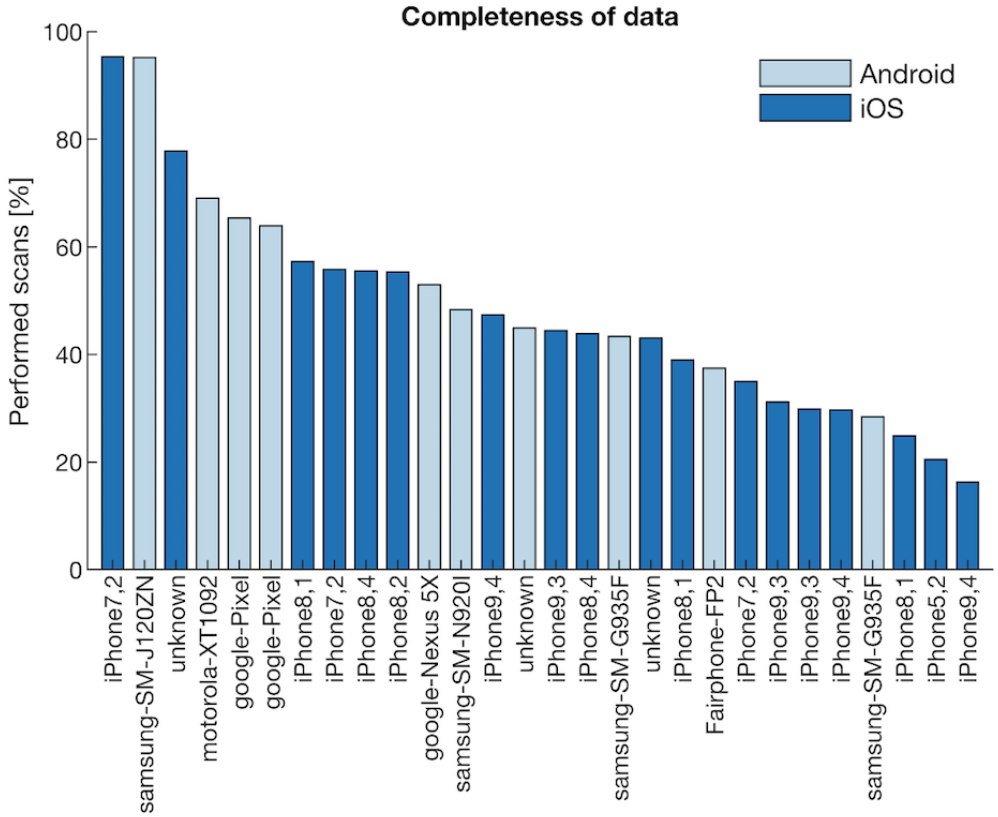


Figure 2. Completeness of data acquired by different devices used by participants.



The scanning rates did not significantly differ between operating systems (t_{26}=−1.33, P=.19, d=0.53). However, the number of scans that were collected varied considerably between devices (range 16.3%-95.4%), approximating a normal distribution (iOS: W=0.93, P=.20; Android: W=0.95, P=.65). We also recorded the model of the device, but there did not appear to be a clear relationship with the scanning rate, as seen in Figure 2.

Passive Data Collection

In this study, we collected two types of sensor data (Bluetooth and GPS) using the Socialise app. Both types of data may provide behavioral indicators of mental health.

Bluetooth Connectivity

When assessing the number of mobile phone devices that were detected using Bluetooth, we observed large variability between participants, both in the total number of devices that were detected and the ratio of known and unknown devices, as seen in the top panel of Figure 3. When considering the average number of nearby mobile phones at different times of the day, few nearby devices were detected during sleeping time (0-6 am), and they were mostly known devices, as seen in the bottom panel of Figure 3. In contrast, office hours had the most device detections and also showed the highest percentage of unknown devices. In the evening, the number of known devices stabilized, whereas the number of unknown devices gradually decreased.

Global Positioning System: Location and Mobility

Location data were captured from participants throughout Australia. The top panel in Figure 4 shows the locations (latitude and longitude) of participants during the 4-week study overlaid on Google maps. Data of individual participants are color coded. The number of location clusters identified for each participant ranged from 4 to 30 with a median of 8 clusters. The bottom panel of Figure 4 shows clusters extracted from a representative participant. Dots represent the centroid of different clusters and the size of dots indicates the number of samples captured within each cluster.

Figure 5 shows the circadian movement measured at scanning intervals of 3 to 8 minutes (displayed as separate lines). Circadian movement measures to what extent the participants' sequence of locations follows a 24-hour rhythm. Lower circadian movement scores indicate that location changes revealed a weaker 24-hour rhythm. A repeated-measures ANOVA showed no significant effect of scanning interval on circadian movement ($F_{3,69}$=2.31, P=.08), indicating that different scanning intervals did not introduce a significant bias in estimating circadian movement. Cronbach alpha was .79 (95% CI 0.61-0.89), indicating an acceptable consistency in the circadian movement estimated at different scanning intervals in different weeks.

Figure 3. Number of Bluetooth devices that were detected. Blue indicates known devices and yellow unknown devices.

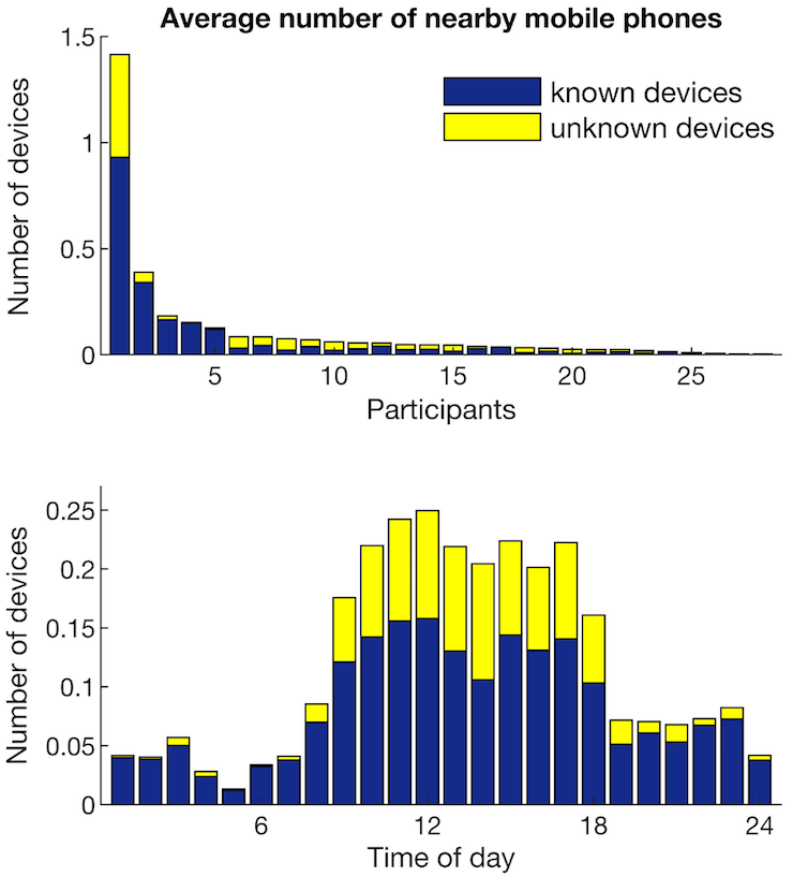


Figure 4. Global positioning system location data of participants during study.

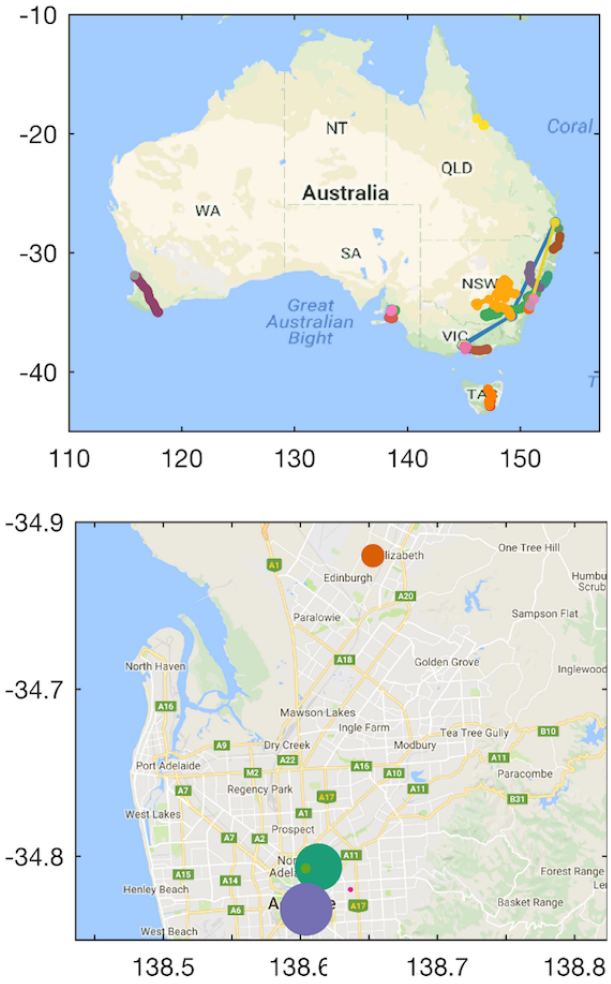
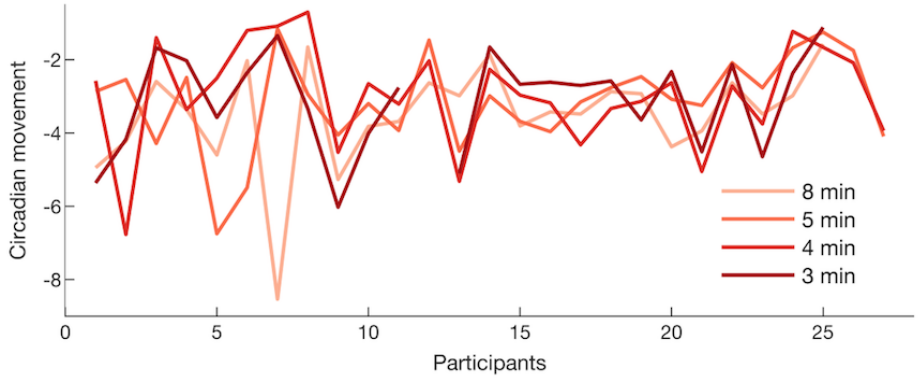


Figure 5. Circadian movement estimated from global positioning system data of individual participants.



User Experience

Battery Consumption

We considered that users typically charge their phones once per day and are awake typically from 6 am to 10 pm (16 hours). With operation of the app, battery life should ideally last at least 16 hours after a full recharge. After systematically varying the time interval between GPS and Bluetooth scans, we used a robust fitting algorithm to estimate the average battery consumption of the Socialise app across devices and scanning rates. Based on the fitted blue regression line seen in Figure 6, we estimated that the average battery life was 21.3 hours when the app did not scan at all and was reduced to 18.8 hours when the app scanned every 5 minutes, resulting in a reduction of 2.5 hours (12%) in battery life. Gray lines show data from individual devices, showing that scanning at 5-minute intervals permitted 5-29 hours of battery life. At this scanning rate, 13 out of 16 devices (81%) had an average battery life of more than 16 hours. At an interscan interval of 3 minutes, average battery life was further reduced to 17.4 hours. In comparison to the small reduction in battery life at increased scanning rates, we observed large variability in battery life across devices, as seen in Figure 6.

Different scanning rates were also subjectively evaluated by asking participants whether they felt that the app impacted their mobile phone's battery life at the end of each week. Participants were asked the question "In the last week, did the app impact the battery life of your phone?" Overall, 23 participants answered the survey question, and 56 ratings were provided over the course of the study. Figure 7 shows the perceived impact of the app on battery life for different scanning frequencies. The percentage of respondents is shown for each of the scores of a 7-point Likert scale, where higher scores indicate greater impact. Colors indicate different scanning rates (once every 8, 5, 4, and 3 minutes with n=18 for 3 and 5 minutes and n=10 for 5 and 8 minutes). The majority of participants reported that battery life was affected by the Socialise app, in particular at higher scanning rates (every 3 or 4 minutes).

Usability

As part of an iterative design and development process, we asked participants to report any problems they experienced in using the Socialise app. Overall, 30 participants (30/32, 94%) answered questions about problems associated with installing and opening the app with half (15/30, 50%) indicating they experienced problems. The most common problem was difficulty logging into the app with the unique participant code (7 participants; Table 2). Many reported problems were technical, which are difficult to address in a preemptive manner because they often depend on user-dependent factors, such as the type, brand, and age of their mobile phones and user behavior (eg, skimming instructions).

Figure 6. Battery life as function of the scanning rate of the Socialise app.

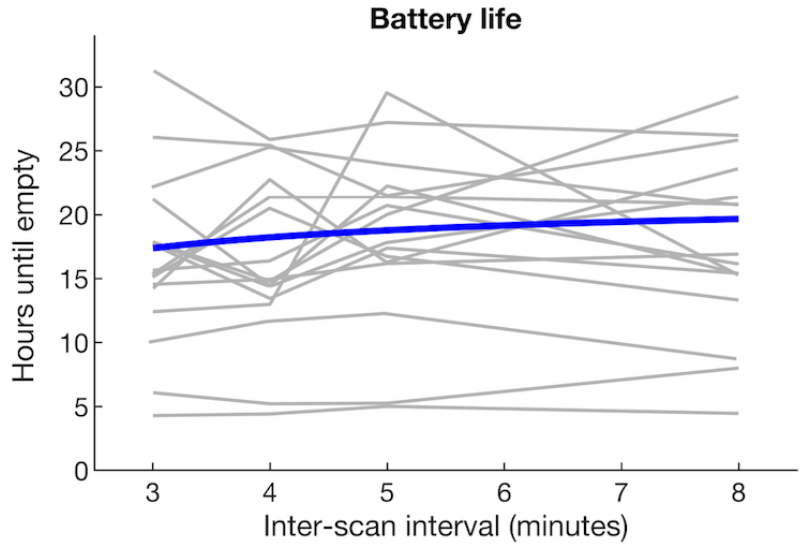


Figure 7. Participant ratings of the impact of the Socialise app on battery life.

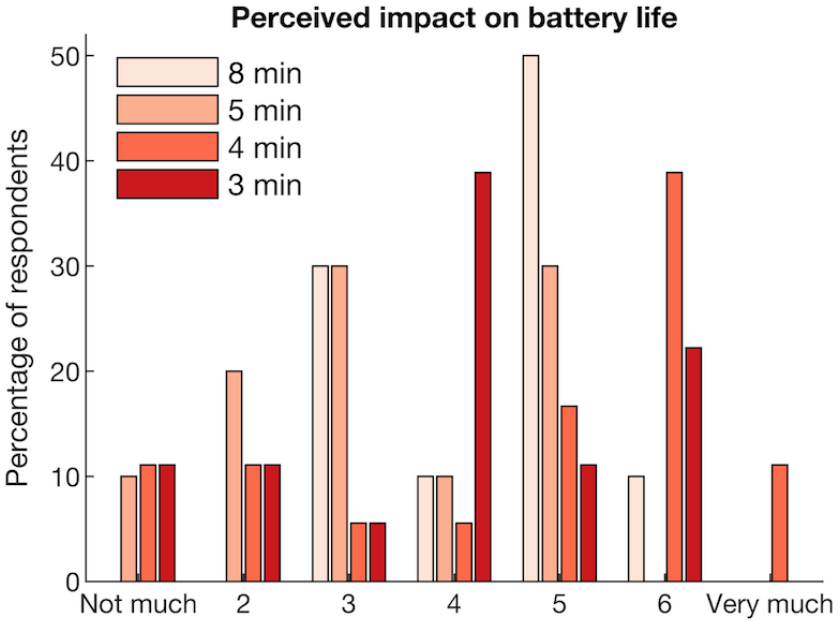


Table 2. Problems experienced installing and opening the Socialise app (n=30).

Response	n (%)	Potential solutions
No problems	15 (50)	—
Problem logging into app	7 (23)	Simplify token
User self-identified lack of proficiency with technology	2 (7)	Improve instructions
App not loading	1 (3)	Improve app release
App needing reinstallation	1 (3)	Improve app release
Phone settings blocking app	1 (3)	Improve instructions
Unspecified problem	3 (10)	—

Table 3. Problems experienced running the Socialise app.

Response	n (%)[a]	Potential solutions
No problem	38 (68)	—
App notification telling users they restarted phones	7 (13)	Only send notification if no data are uploaded to database
Noticeable battery loss	2 (4)	Reduce scanning rate
Difficulty sending emails after app installation	2 (4)	—
App not presenting questionnaires	1 (2)	Check scheduling function of Socialise app
App not scanning	1 (2)	Check settings
Annoying to keep app open; accidentally swipe closed	1 (2)	Send notification if no data are uploaded to database
Unsure if app running properly	1 (2)	Improve instructions
Unsure about what they should be doing with app	1 (2)	Improve instructions
App not working	1 (2)	Check phone model and operating system
Unspecified problem	1 (2)	—

[a]Participants were asked to answer questions about problems running the app four times during the study. Twenty-three unique participants answered these questions, yielding 56 responses.

Fewer participants (23/32, 72%) answered questions about problems they experienced while running the app; these questions were administered at the end of each week. In total, questions were answered 56 times over the course of the study. Just under half (11/23, 48%) of the respondents reported problems running the app, and a problem was identified 32% (18/56) of the time (Table 3). The most common problem was that the app provided a notification to participants stating that they had restarted their phone when users, in fact, had not (7 times). Again, it is evident that a number of encountered problems were technical and, as before, they may be due to mobile phone and user behavior-related factors.

Ethics

To explore ethics and privacy considerations of passive mobile phone sensor data collection, we included a set of survey questions about the acceptability of sensor data collection and some contextual information about that acceptability. Survey questions were administered at the end of the feasibility study (n=13) using a 5-point Likert scale. The top panel of Figure 8 shows that most participants expressed comfort with all aspects of data collection; 77% (10/13) of the participants were either comfortable or very comfortable with GPS, 53% (7/13) with Bluetooth, and 100% (9/9) with questionnaires. A repeated-measures ANOVA showed no main effect of data type ($F_{2,24}$=2.09, P=.15, n=13). We also asked participants how comfortable they were with data collection in different contexts, as seen in the bottom panel of Figure 8. Repeated-measures ANOVA showed a main effect of context ($F_{2.4,29.2}$=7.48, P=.01). Post hoc t tests showed that participants were more comfortable with data collection for research than for advertising (t_{12}=−3.99, P=.002) and for medical intervention than for advertising (t_{12}=3.89, P=.003).

Thematic analysis of responses to open questions revealed the following 3 main themes: uncertainty around the purpose of data collection, helpfulness of data donation to a respected research institute with a secondary theme of trust, and the personal impact of using the app including a secondary theme of perceived impact on privacy.

Participant 11 (henceforth P11), who said they were “Neither comfortable nor uncomfortable” with GPS data collection, explained that “[I was] ok; however, as I was not fully aware of the intentions of the collection of the GPS data and my battery life declining, I started to then get uncomfortable.” Another participant, who also said “neither” for both Bluetooth and GPS tracking said, “I wasn’t sure what the purpose was,” and “[I] don’t understand the implications of this at all” [P12]. P13 said, “Why collect this data?” and “[I] cannot see what value it would be other than to satisfy arbitrary research goals” and felt it to

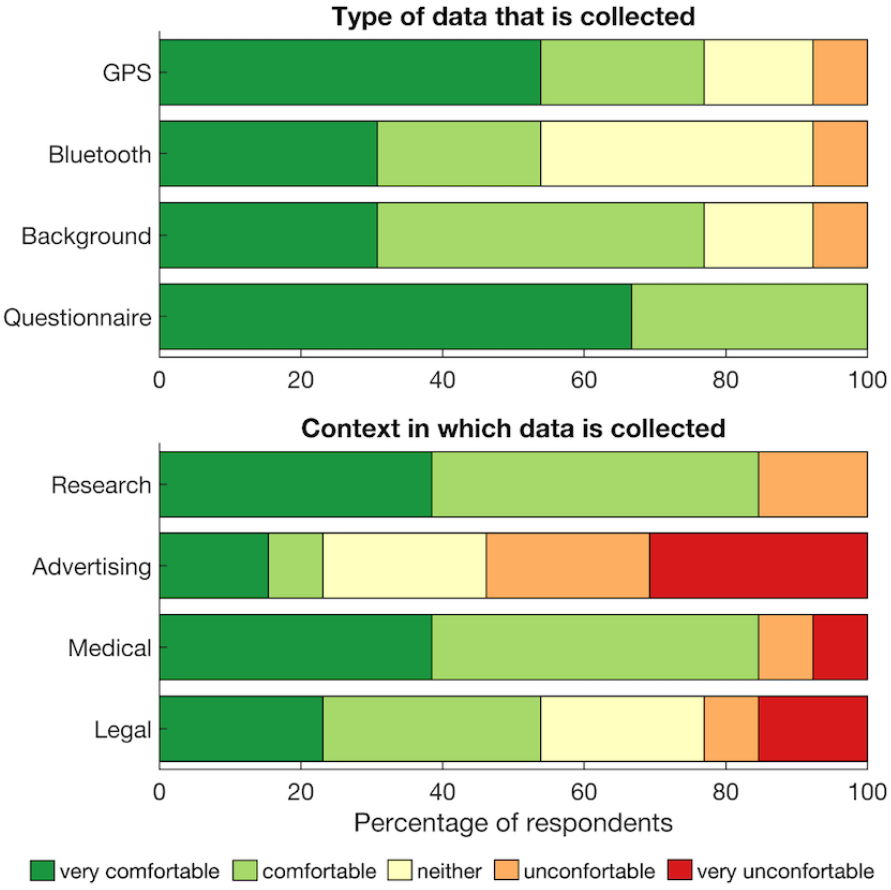
be “an invasion of my privacy.” These responses imply that although the level of discomfort was low overall, a degree of uncertainty existed around the purpose of data collection, and this uncertainty increased discomfort.

Another theme related to the motivation of being helpful to the research or the Institute by providing data. Overall, 4 of the 13 respondents mentioned being helpful as a motivation. P3 was “very comfortable” with GPS tracking and said, “[I] wanted to help in some way.” P2 was quite comfortable with the app running in the background “because I realize that information will be used for the betterment of [the] community.” P7 said, “[I] would like to do anything I can that might help more study,” and P8 would continue using the app or “anything that could help.” This theme is unsurprising given that these users are on a volunteer research register. A second and related theme was around trust. One user explained, “[I] trust the Black Dog [Institute]” (P3) and was therefore comfortable with passive data collection.

Many participants framed their level of comfort with data collection in terms of its perceived effect or impact on them. One participant was “very comfortable” with GPS tracking because “it didn’t affect me” (P4). Others said, “[it] does not bother me” (P2), “[it] did not bother me” (P10), or “[I] did not think much about it” (P9). However, another user who said, “[I was] comfortable” with GPS data collection, explained: “I actually forgot most of the time that it was collecting it. Which slightly made me uncomfortable just in regard to how easily it can happen” (P5). P11, who answered “neither” for effect or impact, said that GPS tracking was impacted by what they felt was draining their battery. P2 also said, “Bluetooth drains battery” and “[I was] uncomfortable” with the Bluetooth being on but also that it was “not a huge problem.” Finally, one user was “uncomfortable” with GPS tracking, explaining, “I believe it is an invasion of my privacy” (P13). However, the same user believed there were “no privacy issues” with Bluetooth data collection.

Another aspect of impact on users was the idea of perceived benefit or lack thereof for them. When responding to a question about whether they would continue to the use the app: “If the app were to be modified showing people you meet and giving information about what it means, I probably would [continue using it]” (P1). However, others said they “don’t see a use for it” (P5) and “[were] not sure how useful it would be for me” (P9). This is not surprising considering that the app is solely for data collection. However, it shows that participants would expect to receive information that they can interpret themselves.

Figure 8. Participants’ comfort with aspects and context of background data collection (GPS: global positioning system).



Discussion

Principle Findings

A feasibility study was conducted to test the Socialise app and examine challenges of using mobile phone sensor technology for mental health research. Sensor data (Bluetooth, GPS, and battery status) was collected for 4 weeks, and views of participants about acceptability of passive sensor technology were investigated. We were able to collect sensor data for about half of the scheduled scans. Social context, location clusters, and circadian movement were features extracted from sensor data to examine behavioral markers that can be obtained using the app. Battery life was reduced by 2.5 hours when scanning every 5 minutes. Despite this limited impact on battery life, most participants reported that running the app noticeably affected their battery life. Participants reported that the purpose of data collection, trust in the organization that collects data, and perceived impact on privacy as important considerations for acceptability of passive data collection.

Behavioral Markers

Instead of assessing social connections between participants, Bluetooth data were used to make a coarse estimate of human density around the participant, which provides a rough proxy for social context. The number and familiarity of devices detected were used to differentiate social contexts. Specifically, more unfamiliar devices were detected during work hours, and fewer familiar devices were detected in the evening. This pattern largely matched that observed by Do et al [47], although the number of overall devices that were detected in our study was lower. This may be partly because we recorded only Bluetooth data from mobile phone devices while filtering out other Bluetooth devices.

We extracted two features from GPS data previously shown to have strong association with self-reported mental health data [7]: circadian movement and location clusters. Circadian movement measures to what extent participants' sequence of locations follows a 24-hour rhythm. Comparing circadian movement assessed separately each week to values across weeks revealed good reliability (Cronbach alpha .79), indicating acceptable consistency in circadian movement estimated in different weeks at different scanning rates. Circadian movement was estimated over 1 week of GPS data, and consistency may be further improved by estimating circadian movement over longer time intervals. We also used a clustering algorithm to identify the number of location clusters that each participant visited. The number of clusters ranged from 4-30 with a median of 8 clusters, which was higher than the number of location clusters reported by Saeb et al [7], ranging from 1-9 with an average of 4.1 clusters. This may be partly due to geographical differences between studies (Australia vs United States). Human mobility patterns are strongly shaped by demographic parameters and geographical contexts, such as age and population density, and it should therefore be determined whether behavioral markers extracted from GPS data are universal or context-dependent [50,51].

Technical Challenges

We were able to collect sensor data for about half of the scheduled scans (Android 55%, iOS 45%). The Socialise app (v0.2) incorporated two technical modifications (ie, using push notifications to trigger scans and using significant-change location service to alert participants when their phone restarted and remind them to resume data collection) to improve data completeness on iOS devices compared with our previous studies, which revealed significant disparity between Android and iOS data acquisition rates using previous versions of the app [43,44]. The 50% data rate in this study is similar to the rate reported in a study using Purple Robot, in which 28 of 40 participants (70%) had data available for more than 50% of the time [7]. However, GPS data of only 18 participants (45%) were used for location analysis in that study, suggesting that the GPS data rate may have been lower. Likewise, in a study using Beiwe in a cohort with schizophrenia, the mean coverage of GPS and accelerometer data were 50% and 47%, respectively [52]. Missing data may limit the number of participants for whom features can be reliably estimated and may also introduce bias in outcome measures extracted from sensor data, for example, participants with fewer data points will appear to have fewer social connections [53]. Interestingly, a recent pilot study (N=16) found that the total coverage of sensor data is itself associated with self-reported clinical symptoms [52].

We found that the Socialise app, when scanning every 5 minutes, reduced battery life from 21.3 hours to 18.8 hours, a 12% reduction. We used silent push notifications to trigger scans intermittently because continuously sampling sensor data would drain the phone's battery in a few hours. Pendão et al [54] estimated that GPS consumed 7% and Bluetooth consumed 4% of total battery power per hour when sampling continuously or 1% and 3%, respectively, when sampling periodically. Therefore, a straightforward solution to conserve battery life is to adjust intervals between data collection points. Longer time intervals between scans and shorter scanning durations can reduce battery consumption, but scanning durations that are too short may not yield meaningful sensor information [23]. Although we used silent push notifications to schedule intermittent scans, other apps use an alternating on-cycle to off-cycle schedule, in which GPS was scheduled to collect data with 1 Hz frequency for a 60-seconds on-cycle, followed by a 600-seconds off-cycle [52]. Another approach to conserve battery is to use conditional sensor activation, for example, adaptive energy allocation [55] and hierarchical sensor management [23]. These solutions reduce the activation of specific sensors at times when they are not needed.

Ethical Considerations

The collection of sensor data involves large quantities of individualized social and behavioral data, and security and privacy have been recognized as a high priority [9,10]. Our participants reported that the purpose of data collection was an important consideration to weigh against any perceived privacy risks, which relates to the theme of uncertainty around purposes of data collection. The consent process for mental health data collection is therefore of importance with regard to both articulating this purpose and outlining confidentiality and risk

of harm to patients [35]. Patient safety should be built into the design of data collection apps. Although this study did not collect mental health data, we intend to use the Socialise app in future studies to assess the mental health symptoms of participants. As such, we have built into the Socialise app a safety alert system, by which participants who indicate high scores on mental health questionnaires will be immediately given contact information about support services and be contacted by a mental health professional to provide additional support. This is consistent with the views of practitioners who have emphasized the importance of including contacts for medical professionals or other services in case of emergency or the need for immediate help [9]. Patients should be made aware of the standard turnaround time for a response to requests for help [2] and administering organizations should ensure that these expectations are clearly defined and consistently met [2].

Our results revealed a degree of uncertainty about the purpose of the study, suggesting that many participants took part without necessarily feeling informed about reasons for it. The communication of purpose should therefore be improved for future studies. Hogle [56] emphasized the need to make a clear distinction whether health-related data are collected for population-level research as opposed to individual, personal treatment or identification of issues. In addition, data processing techniques are often opaque to users, and informed consent may thus be difficult to achieve [42]. Respondents also emphasized their willingness to help the organization with its research and their trust in the organization as a stand-in for certainty about how data would be used. We believe that researchers should not rely on organizational trust as a stand-in for true understanding and informed consent because there is a risk of breach of trust if data are not used as expected.

Other issues included data ownership and the direction of any benefits created, considering that the data are from users [40]. Pentland et al [57] argued that participants should have ownership over their own data, by which they mean that app users should maintain the rights of possession, use, and disposal with some limitations on the right to disclose data about others in one's network. This can be achieved by holding users' data much as a bank would, with informed consent, or by storing data locally on a user's device and requiring upload for analysis [57]. However, when it comes to data, it is those with the capacity to store, analyze, and transfer data who have meaningful power over it; therefore, the concept of data ownership is limited [58].

Passive sensor data may be used for predictive analytics to identify those at risk of mental health issues. However, there is a possibility that predictive models may increase inequalities for vulnerable groups [40], particularly when commercial interests are at play. Psychiatric profiling will identify some as being at high risk, which may shape self-perception [59] and beliefs about an individual. This is particularly significant if the individual is a minor [2]. Hence, nonmedical and commercial use of this data to estimate mental state and behavior is an area of concern [2].

Recommendations

Based on these findings and the literature on passive sensing, usability, and ethics, we make the following recommendations for future research on passive sensing in mental health.

Reporting of Data Completeness and Battery Consumption to Benchmark Different Technical Solutions

Standard reporting of meta-data will enable benchmarking of apps and identification of technical obstacles and solutions for sensor data collection across devices and operating systems. For example, we estimated that the Socialise app reduced battery life by 2.5 hours when scanning every 5 minutes. Although the app had small effect on battery consumption (81% of devices had an average battery life of more than 16 hours), users were very sensitive to battery performance. Standard reporting of data rates and battery consumption will allow quantitative comparisons between approaches and develop technical solutions that meet user expectations on battery life.

Releasing Source Code of Data Acquisition Platforms and Feature Extraction Methods

The number of mobile phone apps for passive sensing is still increasing, but differences in methodology and feature extraction methods can impede the reproducibility of findings. This can be overcome with a commitment to open science because a number of elements of passive data research could be shared. Currently, several sensing platforms are open source, such as Purple Robot [6] and recently, Beiwe [52]. Following this lead, methods for feature extraction could be made open source, such that scripts are available for use on different data sources, providing consistency in feature extraction. Finally, the data itself should be made available on open data repositories to enable data aggregation across studies to test potential markers in larger samples, resulting in more reproducible results [60]. However, data sharing not only has great potential but also involves concerns about privacy, confidentiality, and control of data on individuals [61]. These concerns particularly apply to sensor data such as GPS that can be reidentified [62]. Databases that allow analysis to be conducted without access to raw data may be one potential solution.

Identifying a Limited Number of Key Markers for Mental Health

Although the use of passive data in mental health is still exploratory, researchers need to move toward agreement on best practice methods and features. The current unrestricted number of features has the danger of inflating degrees of freedom and may endanger replicability of findings [63]. Practices such as preregistration of study hypotheses and proposed methods to quantify features could help reduce spurious correlations and will be key in identifying reliable markers of mental health [64]. However, work with different sensor modalities is at different stages of development. For example, a number of GPS features have been identified and replicated [6], whereas potential markers of social connectedness using Bluetooth data still require research to assess predictive value in open network settings. This development of new methods of data analysis is indeed one of the most immediate

challenges [5]. Once candidate methods have been identified, it will be important to test these markers in larger longitudinal studies to see whether they predict the development of mental health problems and can be used to support prevention and early intervention programs [65].

Providing Meaningful Feedback to Users

User engagement is also a key requirement for successful implementation of sensor technology in mental health research. Investigating user experience can help us understand user expectations and improve user engagement and retention [66]. Although passive data collection is designed to be unobtrusive, perceived benefit is an important consideration for continued use of mental health apps. A user-centric design process [27] and the American Psychiatric Association's app evaluation model [67] should be followed to provide meaningful user feedback from sensor data. We also recommend using more robust measures for informed consent, considering the opacity of data analysis techniques and purposes [47] and engaging users with informative feedback derived from their data.

Transparency in the Purpose of Data Collection

Evidence from the literature and participant responses suggests that purposes of data collection are important as well as the awareness of the user. The use of data was found to be most the important factor in a person's willingness to share their electronic personal health data [10], and participants cared most about the specific purpose for using their health information [68]. Rothstein argued that there is too much emphasis on privacy when the concern should be about autonomy [69]. This refers to the informed consent process, during which researchers should ensure understanding and enable autonomous and active consent on that basis [69]. It is therefore recommended that researchers take care to ensure that the consent process allows participants really to understand the purpose of the research. This, in turn, is likely to increase the level of comfort with data collection.

Conclusion

The use of passive data in mental health research has the potential to change the nature of identification and treatment of mental health disorders. Early identification of behavioral markers of mental health problems will allow us to preempt rather than respond, and understanding idiosyncratic patterns will enable personalized dynamic treatment delivered at the moment. Although a number of significant technological and broader challenges exist, we believe that open science, user involvement, collaborative partnerships, and transparency in our attempts, successes, and failures will bring us closer to this goal.

Data Availability

Data used in this study is available at Zenodo: http://doi.org/10.5281/zenodo.1238226. One participant did not consent to have individual data made publicly available. We did not share GPS data because this would allow reidentification of participants. The Matlab scripts used to analyze data are available at Zenodo: http://doi.org/10.5281/zenodo.1238408.

Acknowledgments

This research was financially supported by the National Health and Medical Research Council (NHMRC) Centre of Research Excellence in Suicide Prevention APP1042580 and NHMRC John Cade Fellowship APP1056964. TWB was supported by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey Questions.

[PDF File (Adobe PDF File), 45KB - jmir_v20i7e10131_app1.pdf]

References

1. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
2. Glenn T, Monteith S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. Curr Psychiatry Rep 2014 Dec;16(12):523. [doi: 10.1007/s11920-014-0523-3] [Medline: 25308392]
3. Torous J, Staples P, Onnela JP. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602. [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]
4. Harari GM, Müller SR, Aung MS, Rentfrow PJ. Smartphone sensing methods for studying behavior in everyday life. Curr Opin Behav Sci 2017 Dec;18:83-90. [doi: 10.1016/j.cobeha.2017.07.018]
5. Onnela JP, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]

6. Schueller SM, Begale M, Penedo FJ, Mohr DC. Purple: a modular system for developing and deploying behavioral intervention technologies. J Med Internet Res 2014;16(7):e181 [FREE Full text] [doi: 10.2196/jmir.3376] [Medline: 25079298]
7. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
8. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
9. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
10. Stopczynski A, Sekara V, Sapiezynski P, Cuttone A, Madsen MM, Larsen JE, et al. Measuring large-scale social networks with high resolution. PLoS One 2014;9(4):e95978 [FREE Full text] [doi: 10.1371/journal.pone.0095978] [Medline: 24770359]
11. Cacioppo JT, Hughes ME, Waite LJ, Hawkley LC, Thisted RA. Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. Psychol Aging 2006 Mar;21(1):140-151. [doi: 10.1037/0882-7974.21.1.140] [Medline: 16594799]
12. Barnett PA, Gotlib IH. Psychosocial functioning and depression: distinguishing among antecedents, concomitants, and consequences. Psychol Bull 1988 Jul;104(1):97-126. [Medline: 3043529]
13. Pachucki MC, Ozer EJ, Barrat A, Cattuto C. Mental health and social networks in early adolescence: a dynamic study of objectively-measured social interaction behaviors. Soc Sci Med 2015 Jan;125:40-50. [doi: 10.1016/j.socscimed.2014.04.015] [Medline: 24797692]
14. Campbell K, Lee B. Name generators in surveys of personal networks. Soc Networks 1991;13(3):203-221.
15. Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci U S A 2009 Sep 8;106(36):15274-15278 [FREE Full text] [doi: 10.1073/pnas.0900282106] [Medline: 19706491]
16. Pachucki MC, Ozer EJ, Barrat A, Cattuto C. Mental health and social networks in early adolescence: a dynamic study of objectively-measured social interaction behaviors. Soc Sci Med 2015 Jan;125:40-50. [doi: 10.1016/j.socscimed.2014.04.015] [Medline: 24797692]
17. Prigerson HG, Monk TH, Reynolds CF, Begley A, Houck PR, Bierhals AJ, et al. Lifestyle regularity and activity level as protective factors against bereavement-related depression in late-life. Depression 1995;3(6):297-302. [doi: 10.1002/depr.3050030607]
18. Roshanaei-Moghaddam B, Katon WJ, Russo J. The longitudinal effects of depression on physical activity. Gen Hosp Psychiatry 2009;31(4):306-315. [doi: 10.1016/j.genhosppsych.2009.04.002] [Medline: 19555789]
19. Teychenne M, Ball K, Salmon J. Physical activity and likelihood of depression in adults: a review. Prev Med 2008 May;46(5):397-411. [doi: 10.1016/j.ypmed.2008.01.009] [Medline: 18289655]
20. Montoliu R, Blom J, Gatica-Perez D. Discovering places of interest in everyday life from smartphone data. Multimed Tools Appl 2012 Jan 10;62(1):179-207. [doi: 10.1007/s11042-011-0982-z]
21. Chen J, Bierlaire M. Probabilistic Multimodal Map Matching With Rich Smartphone Data. Journal of Intelligent Transportation Systems 2014 Sep 29;19(2):134-148. [doi: 10.1080/15472450.2013.764796]
22. Wikipedia. Mobile phone based sensing software. URL: https://en.wikipedia.org/wiki/Mobile_phone_based_sensing_software [accessed 2018-02-14] [WebCite Cache ID 6xDag55nm]
23. Wang Y, Lin J, Annavaram M, Jacobson Q, Hong J, Krishnamachari B. A framework of energy efficient mobile sensing for automatic user state recognition. 2009 Presented at: Proceedings of the 7th international conference on Mobile systems, applications, and services; June 22-25, 2009; Kraków, Poland. [doi: 10.1145/1555816.1555835]
24. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
25. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
26. Arsand E, Demiris G. User-centered methods for designing patient-centric self-help tools. Inform Health Soc Care 2008 Sep;33(3):158-169. [doi: 10.1080/17538150802457562] [Medline: 18850399]
27. 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;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
28. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]

29. 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]
30. 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]
31. Bardram J, Frost M, Szántó K, Marcu G, editors. The MONARCA self-assessment system: a persuasive personal monitoring system for bipolar patients. 2012 Presented at: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium; January 28-30, 2012; Miami, Florida, USA. [doi: 10.1145/2110363.2110370]
32. de MY, Hidalgo CA, Verleysen M, Blondel VD. Unique in the Crowd: The privacy bounds of human mobility. Sci Rep 2013;3:1376 [FREE Full text] [doi: 10.1038/srep01376] [Medline: 23524645]
33. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854. [doi: 10.1177/1745691616650285] [Medline: 27899727]
34. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. JMIR Mhealth Uhealth 2015;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
35. Clemens NA. Privacy, consent, and the electronic mental health record: The Person vs. the System. J Psychiatr Pract 2012 Jan;18(1):46-50. [doi: 10.1097/01.pra.0000410987.38723.47] [Medline: 22261983]
36. Jones N, Moffitt M. Ethical guidelines for mobile app development within health and mental health fields. Professional Psychology: Research and Practice 2016;47(2):155-162. [doi: 10.1037/pro0000069]
37. Proudfoot J, Parker G, Hadzi PD, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
38. Shaw H, Ellis DA, Kendrick L, Ziegler F, Wiseman R. Predicting Smartphone Operating System from Personality and Individual Differences. Cyberpsychol Behav Soc Netw 2016 Dec;19(12):727-732. [doi: 10.1089/cyber.2016.0324] [Medline: 27849366]
39. Reinfelder L, Benenson Z, Gassmann F. Differences between Android and iPhone users in their security and privacy awareness. In: SpringerLink.: Springer, Cham; 2014 Presented at: International Conference on Trust, Privacy and Security in Digital Business; 2014; Lyon, France. [doi: 10.1007/978-3-319-09770-1_14]
40. Monteith S, Glenn T, Geddes J, Whybrow PC, Bauer M. Big data for bipolar disorder. Int J Bipolar Disord 2016 Dec;4(1):10 [FREE Full text] [doi: 10.1186/s40345-016-0051-7] [Medline: 27068058]
41. Grande D, Mitra N, Shah A, Wan F, Asch DA. Public preferences about secondary uses of electronic health information. JAMA Intern Med 2013 Oct 28;173(19):1798-1806 [FREE Full text] [doi: 10.1001/jamainternmed.2013.9166] [Medline: 23958803]
42. Andrejevic M. The big data divide. Int J Commun. ISSN 2014;8:17.
43. Boonstra TW, Werner-Seidler A, O'Dea B, Larsen ME, Christensen H. Smartphone app to investigate the relationship between social connectivity and mental health. Conf Proc IEEE Eng Med Biol Soc 2017 Dec;2017:287-290. [doi: 10.1109/EMBC.2017.8036818] [Medline: 29059866]
44. Boonstra T, Larsen M, Townsend S, Christensen H. Validation of a smartphone app to map social networks of proximity. PLoS One 2017;12(12):e0189877. [Medline: 29261782]
45. Townsend S, Larsen M, Boonstra T, Christensen H. Using Bluetooth Low Energy in smartphones to map social networks. arXiv 2015:150803938.
46. Do TMT, Gatica-Perez D. Human interaction discovery in smartphone proximity networks. Pers Ubiquit Comput 2011 Dec 14;17(3):413-431. [doi: 10.1007/s00779-011-0489-7]
47. Do T, Blom J, Gatica-Perez D. Smartphone usage in the wild: a large-scale analysis of applications and context. 2011 Presented at: Proceedings of the 13th international conference on multimodal interfaces; 2011; Alicante, Spain. [doi: 10.1145/2070481.2070550]
48. Press W. Numerical recipes 3rd edition: The art of scientific computing. Cambridge, UK: Cambridge University Press; 2007.
49. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association 1979 Dec;74(368):829-836. [doi: 10.1080/01621459.1979.10481038] [Medline: 26281194]
50. Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C. A tale of many cities: universal patterns in human urban mobility. PloS One 2012;7(5):e37027. [Medline: 22666339]
51. Song C, Qu Z, Blumm N, Barabási AL. Limits of predictability in human mobility. Science 2010 Feb 19;327(5968):1018-1021 [FREE Full text] [doi: 10.1126/science.1177170] [Medline: 20167789]
52. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. npj Digital Med 2018 Apr 6;1(1):15. [doi: 10.1038/s41746-018-0022-8] [Medline: 28688902]

53. Boonstra T, Larsen M, Christensen H. Mapping dynamic social networks in real life using participants' own smartphones. Heliyon 2015;1(3):e00037. [Medline: 27441223]
54. Pendão C, Moreira A, Rodrigues H. Energy consumption in personal mobile devices sensing applications. In: Wireless and Mobile Networking Conference (WMNC). 2014 Presented at: 2014 7th IFIP Wireless and Mobile Networking Conference (WMNC); 20-22 May 2014; Vilamoura, Portugal. [doi: 10.1109/WMNC.2014.6878874]
55. Lin M, Lane N, Mohammod M, Yang X, Lu H, Cardone G. BeWell+: multi-dimensional wellbeing monitoring with community-guided user feedback and energy optimization. 2012 Presented at: Proceedings of the conference on Wireless Health; October 23-25, 2012; San Diego, California. [doi: 10.1145/2448096.2448106]
56. Hogle LF. Data-intensive resourcing in healthcare. BioSocieties 2016;11(3):372-393.
57. Pentland A, Lazer D, Brewer D, Heibeck T. Using reality mining to improve public health and medicine. Stud Health Technol Inform 2009;149:93-102. [Medline: 19745474]
58. Wigan MR, Clarke R. Big Data's Big Unintended Consequences. Computer 2013 Jun;46(6):46-53. [doi: 10.1109/MC.2013.195] [Medline: 27295638]
59. Mehrotra A, Hendley R, Musolesi M, editors. Towards multi-modal anticipatory monitoring of depressive states through the analysis of human-smartphone interaction. : Adjunct; 2016 Presented at: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12 - 16, 2016; Heidelberg, Germany. [doi: 10.1145/2968219.2968299]
60. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open research culture. Science 2015 Jun 26;348(6242):1422-1425 [FREE Full text] [doi: 10.1126/science.aab2374] [Medline: 26113702]
61. Kostkova P, Brewer H, de LS, Fottrell E, Goldacre B, Hart G, et al. Who Owns the Data? Open Data for Healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
62. Green B, Cunningham G, Ekblaw A, Kominers P. Open Data Privacy. SSRN Journal 2017:111. [doi: 10.2139/ssrn.2924751]
63. Simmons J, Nelson L, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci 2011;22(11):1359-1366. [Medline: 22006061]
64. Nosek BA, Lakens D. Registered Reports. Social Psychology 2014 May;45(3):137-141. [doi: 10.1027/1864-9335/a000192]
65. Calear AL, Christensen H. Systematic review of school-based prevention and early intervention programs for depression. J Adolesc 2010 Jun;33(3):429-438. [doi: 10.1016/j.adolescence.2009.07.004] [Medline: 19647310]
66. Lemon C. Psychiatric Times. The user experience: A key step in realizing the role of mental health apps URL: http://www.psychiatrictimes.com/schizophrenia/user-experience-key-step-realizing-role-mental-health-apps [accessed 2018-06-13] [WebCite Cache ID 708ExkDMv]
67. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen T, Luo J, et al. A Hierarchical Framework for Evaluation and Informed Decision Making Regarding Smartphone Apps for Clinical Care. Psychiatr Serv 2018 Feb 15:appips201700423. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
68. Larson EB. Building trust in the power of "big data" research to serve the public good. JAMA 2013 Jun 19;309(23):2443-2444. [doi: 10.1001/jama.2013.5914] [Medline: 23780455]
69. Rothstein M. Ethical issues in big data health research: currents in contemporary bioethics. J Law Med Ethics 2015;43(2):425-429. [Medline: 26242964]

Abbreviations

ANOVA: analysis of variance
EMA: ecological momentary assessment
GPS: global positioning system
NHMRC: National Health and Medical Research Council
PTSD: Posttraumatic Stress Disorder

Edited by R Calvo, M Czerwinski, J Torous, G Wadley; submitted 16.02.18; peer-reviewed by P Chow, IW Pulantara, I Mohino-Herranz; comments to author 01.04.18; revised version received 22.05.18; accepted 12.06.18; published 30.07.18.

Please cite as:
Boonstra TW, Nicholas J, Wong QJJ, Shaw F, Townsend S, Christensen H
Using Mobile Phone Sensor Technology for Mental Health Research: Integrated Analysis to Identify Hidden Challenges and Potential Solutions
J Med Internet Res 2018;20(7):e10131
URL: http://www.jmir.org/2018/7/e10131/
doi:10.2196/10131
PMID:30061092

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org