Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 19 (2017), Issue 11 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Assessing the Efficacy of Mobile Health Apps Using the Basic Principles of Cognitive Behavioral Therapy: Systematic Review (e399)
Amy Rathbone, Laura Clarry, Julie Prescott. . . . 4

Self-Guided Web-Based Interventions: Scoping Review on User Needs and the Potential of Embodied Conversational Agents to Address Them (e383)
Mark Scholten, Saskia Kelders, Julia Van Gemert-Pijnen. . . . 126

Patient Health Record Systems Scope and Functionalities: Literature Review and Future Directions (e388)
Lina Bouayad, Anna Ialynytchev, Balaji Padmanabhan. . . . 318

Viewpoint

Is Connected Health Contributing to a Healthier Population? (e386)
Carmen Loiselle, Saima Ahmed. . . . 14

Original Papers

Engagement Within a Mobile Phone–Based Smoking Cessation Intervention for Adolescents and its Association With Participant Characteristics and Outcomes (e356)
Raquel Paz Castro, Severin Haug, Andreas Filler, Tobias Kowatsch, Michael Schaub. . . . 18

Preventing Depression in Final Year Secondary Students: School-Based Randomized Controlled Trial (e369)
Yael Perry, Aliza Werner-Seidler, Alison Calear, Andrew Mackinnon, Catherine King, Jan Scott, Sally Merry, Theresa Fleming, Karolina Stasiak, Helen Christensen, Philip Batterham. . . . 33

Sexual Health and the Internet: Cross-Sectional Study of Online Preferences Among Adolescents (e379)
Antonella von Rosen, Frederik von Rosen, Peter Tinnemann, Falk Müller-Riemenschneider. . . . 44

Reducing Symptom Distress in Patients With Advanced Cancer Using an e-Alert System for Caregivers: Pooled Analysis of Two Randomized Clinical Trials (e354)
David Gustafson, Lori DuBenske, Amy Atwood, Ming-Yuan Chih, Roberta Johnson, Fiona McTavish, Andrew Quanbeck, Roger Brown, James Cleary, Dhavan Shah. . . . 54

A Fully Automated Web-Based Program Improves Lifestyle Habits and HbA1c in Patients With Type 2 Diabetes and Abdominal Obesity: Randomized Trial of Patient E-Coaching Nutritional Support (The ANODE Study) (e360)
Boris Hansel, Philippe Giral, Laetitia Gambotti, Alexandre Lafourcade, Gilbert Peres, Claude Filipecki, Diana Kadouch, Agnes Hartemann, Jean-Michel Oppert, Eric Bruckert, Michel Marre, Arnaud Bruneel, Emilie Duchene, Ronan Roussel. 65

Brain Activation in Response to Personalized Behavioral and Physiological Feedback From Self-Monitoring Technology: Pilot Study (e384)
Maxine Whelan, Paul Morgan, Lauren Sherar, Andrew Kingsnorth, Daniele Magistro, Dale Esliger. 79

Tailored Web-Based Interventions for Pain: Systematic Review and Meta-Analysis (e385)
Geraldine Martorella, Madalina Boitor, Melanie Berube, Suzanne Fredericks, Sylvie Le May, Céline Gélinas. 94

Effectiveness of a Web 2.0 Intervention to Increase Physical Activity in Real-World Settings: Randomized Ecological Trial (e390)
Corneel Vandelanotte, Gregory Kolt, Cristina Caperchione, Trevor Savage, Richard Rosenkranz, Anthony Maeder, Anetta Van Itallie, Rhys Tague, Christopher Oldmeadow, W Mummery, Mitch Duncan. 111

Reduction in Vegetable Intake Disparities With a Web-Based Nutrition Education Intervention Among Lower-Income Adults in Japan: Randomized Controlled Trial (e377)
Saki Nakamura, Takayo Inayama, Kazuhiro Harada, Takashi Arao. 145

Implementation of a Web-Based Organ Donation Educational Intervention: Development and Use of a Refined Process Evaluation Model (e396)
Nakeva Redmond, Laura Harker, Yvan Bamps, Shauna Flemming, Jennie Perryman, Nancy Thompson, Rachel Patzer, Nancy Williams, Kimberly Arriola. 163

Relationship Between the Menstrual Cycle and Timing of Ovulation Revealed by New Protocols: Analysis of Data from a Self-Tracking Health App (e391)
Satoshi Sohda, Kenta Suzuki, Ichiro Igari. 177

Implementation Intention and Reminder Effects on Behavior Change in a Mobile Health System: A Predictive Cognitive Model (e397)
Peter Pirolli, Shiwali Mohan, Anusha Venkatakrishnan, Les Nelson, Michael Silva, Aaron Springer. 188

Health-Seeking Influence Reflected by Online Health-Related Messages Received on Social Media: Cross-Sectional Survey (e382)
Rahila Iftikhar, Bahaa Abaalkhail. 205

Using Facebook to Recruit Young Australian Men Into a Cross-Sectional Human Papillomavirus Study (e389)
Roopa Das, Dorothy Machalek, Edmund Molesworth, Suzanne Garland. 217

Social Media Interventions to Promote HIV Testing, Linkage, Adherence, and Retention: Systematic Review and Meta-Analysis (e394)
Bolin Cao, Somya Gupta, Jiangtao Wang, Lisa Hightow-Weidman, Kathryn Muessig, Weiming Tang, Stephen Pan, Razia Pendse, Joseph Tucker. 227

Two Novel Cognitive Behavioral Therapy–Based Mobile Apps for Agoraphobia: Randomized Controlled Trial (e398)
Marina Christoforou, José Sáez Fonseca, Elias Tsakanikos. 240

Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial (e373)
Akio Mantani, Tadashi Kato, Toshi Furukawa, Masaru Horikoshi, Hissei Imai, Takahiro Hiroe, Bun Chino, Tadashi Funayama, Naohiro Yonemoto, Qi Zhou, Nao Kawanishi. . . . 253

Subregional Nowcasts of Seasonal Influenza Using Search Trends (e370)
Sasikiran Kandula, Daniel Hsu, Jeffrey Shaman. . . . 269

Understanding Older Adult's Technology Adoption and Withdrawal for Elderly Care and Education: Mixed Method Analysis from National Survey (e374)
Ching-Ju Chiu, Chia-Wen Liu. . . . 285

Bioimpedance Alerts from Cardiovascular Implantable Electronic Devices: Observational Study of Diagnostic Relevance and Clinical Outcomes (e393)
Christophe Smeets, Julie Vranken, Jo Van der Auwera, Frederik Verbrugge, Wilfried Mullens, Matthias Dupont, Lars Grieten, Hélène De Cannière, Dorien Lanssens, Thijs Vandenberk, Valerie Storms, Inge Thijs, Pieter Vandervoort. . . . 296

Development and Validation of a Taxonomy for Characterizing Measurements in Health Self-Quantification (e378)
Manal Almalki, Kathleen Gray, Fernando Martin-Sanchez. . . . 307

Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques (e395)
Jim Lumsden, Andy Skinner, David Coyle, Natalia Lawrence, Marcus Munafo. . . . 333

Performance of a Mobile Phone App-Based Participatory Syndromic Surveillance System for Acute Febrile Illness and Acute Gastroenteritis in Rural Guatemala (e368)
Daniel Olson, Molly Lamb, Maria Lopez, Kathryn Colborn, Alejandra Paniagua-Avila, Alma Zacarias, Ricardo Zambrano-Perilla, Sergio Rodríguez-Castro, Celia Cordon-Rosales, Edwin Asturias. . . . 351

Artificial Intelligence Learning Semantics via External Resources for Classifying Diagnosis Codes in Discharge Notes (e380)
Chin Lin, Chia-Jung Hsu, Yu-Sheng Lou, Shih-Jen Yeh, Chia-Cheng Lee, Sui-Lung Su, Hsiang-Cheng Chen. . . . 362

Public Awareness and Use of German Physician Ratings Websites: Cross-Sectional Survey of Four North German Cities (e387)
Stuart McLennan, Daniel Strech, Andrea Meyer, Hannes Kahrass. . . . 375

Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies (e367)
Trisha Greenhalgh, Joseph Wherton, Chrysanthi Papoutsi, Jennifer Lynch, Gemma Hughes, Christine A'Court, Susan Hinder, Nick Fahy, Rob Procter, Sara Shaw. . . . 386

Review

Assessing the Efficacy of Mobile Health Apps Using the Basic Principles of Cognitive Behavioral Therapy: Systematic Review

Amy Leigh Rathbone[1], BSc (Hons); Laura Clarry[1], BSc (Hons); Julie Prescott[1], PhD

Department of Education and Psychology, University of Bolton, Bolton, United Kingdom

Corresponding Author:
Julie Prescott, PhD
Department of Education and Psychology
University of Bolton
Deane Road
Bolton, BL3 5AB
United Kingdom
Phone: 44 1204903676
Email: j.prescott@bolton.ac.uk

Abstract

Background: Cognitive behavioral therapy (CBT) in its basic principle has developed itself as a stand-alone, substantial method of therapy. With effective application in therapy for a range of mental health issues, the spread of CBT methods to Web-based therapy sources is evident. The development of mobile phone apps using CBT principles is increasing within the research area. Despite the move to Web-based methods of therapy, it is argued that these methods lack the same efficacy of face-to-face therapy sessions.

Objective: The aim of this review was to assess extent research findings with regard to the effectiveness of CBT-related mobile health (mHealth) apps. By assessing only studies employing a randomized controlled trial design, the review aimed to determine app efficacy within the highly regarded method of investigation.

Methods: A comprehensive literature search was conducted across several databases. Search results were filtered, and results were subject to strict inclusion and exclusion criteria because of the nature of the review. Where possible, analysis of effect size was calculated and results reported.

Results: A total of 8 studies investigating the effectiveness of mHealth CBT-related apps across a range of mental health issues were reviewed. Three studies used the app against a control group, and 5 studies used the app intervention against another form of treatment or intervention. A range of effect sizes were seen across all included studies (d=−0.13 to 1.83; 0.03-1.44), with the largest effects often being seen when comparing the data from pre- to posttest for the app engaged group.

Conclusions: The studies reviewed support the use of mHealth apps containing CBT principles for a range of mental health issues. However, the effectiveness over longer time periods should be assessed. Researchers and professionals should seek to collaborate effectively when creating new apps to enhance their effectiveness as a treatment for the general public.

(J Med Internet Res 2017;19(11):e399) doi:10.2196/jmir.8598

KEYWORDS

cognitive therapy; behavior; mHealth; mobile; phone; health; treatment efficacy; intervention study; randomized controlled trial; review; systematic

Introduction

Cognitive Behavioral Therapy

Cognitive behavioral therapy (CBT) is a substantially adapted, personalized, psychosocial therapy [1]. It has emerged as a viable, empirically reinforced treatment for various mental health issues [2]. The therapy proposes maladaptive cognition correlates to a cognitive and attention bias toward misinterpretation of information and perceiving certain information to be directly threatening and cataclysmic [3]. CBT focuses on personal connotations added to situations and produces empowering psychological strategies to reevaluate the meanings attributed to situations, promoting learned practice for more positively altered behaviors, emotions, and thoughts [4].

Beck et al [5] theorized that, "Cognitions (verbal or pictorial 'events' in a person's stream of consciousness) are based on attitudes or assumptions (schemas), developed from previous experiences."

Beck et al [6] went on to add, "The psychological sequence progresses from evaluation to affective and motivational arousal, and finally to selection, and implementation of a relevant strategy."

These two short quotes encompass the basic principles of CBT. CBT consists of three core principles: cognitive activity affects behavior, cognitive activity may be monitored and altered, and desired behavioral change can occur through cognitive change [7]. The therapy aims to target negative emotions that can not only be overwhelming but can also have a detrimental effect on a person's quality of life. Negative emotions, when experienced in the correct context, can be a typical occurrence, for example, stress, bereavement, anger, or jealousy. However, if these feelings increase and occur exponentially, physical symptoms such as increased blood pressure, headaches, insomnia, and loss of libido can develop.

The therapy itself is tailored to meet the patients' needs and aims to utilize the previously formed therapeutic alliance between patient and therapist to recognize and comprehend present difficulties [8]. Patients are required to engage with *homework* outside of therapeutic hours. This encourages patients to foster a sense of cognitive self-awareness away from a clinical setting.

Applications and Efficacy

CBT can be applied to a variety of mental health issues and when utilized in the proper manner, or comorbid with other relevant treatment, result in a significant reduction in symptomology. One such issue is obsessive compulsive disorder (OCD).

Recent research has found that CBT has the potential to reduce and control symptomology of OCD in a way that far exceeds the pharmacological methods such as serotonergic antidepressants [9]. Several studies found that CBT garnered a greater effect size and provided more substantial improvements of clinical symptoms [10-13]. A meta-analysis carried out by Olatunji et al [14] examined the efficacy of CBT for OCD and found that the therapy was highly effective for the reduction of symptomology. The study found that there were large effect sizes immediately post treatment and medium effect sizes during a follow-up.

CBT has shown effectiveness for treating both children [15] and adults [16] suffering from posttraumatic stress disorder (PTSD). CBT specifically used for PTSD tends to be trauma-focused. Trauma-focused CBT (TF-CBT) is a highly versatile model of psychotherapy that focuses on trauma-specific emotional stimulus [17]. When implementing CBT for PTSD in a randomized controlled trial (RCT), Smith et al [18] found that after a month-long symptom observation period, 24% of young people with a preliminary diagnosis of PTSD developed so much that they actually failed to meet the criterion for the disorder. The same study also found that TF-CBT reduced symptoms of not only PTSD but depression and anxiety and also enabled a better quality of life. These results were also maintained over a span of 6 months.

Previous studies and meta-analyses have shown that CBT is an efficacious therapy for anxiety disorders. There have been results showing that CBT has the benefit of reducing comorbidity of multiple anxiety disorders [19]; its clinical utility is valid when treating anxious children and avoiding relapse in adolescents and adulthood [20] and overall reduction of anxiety disorder symptomology [21].

CBT has shown positive results in many other areas of mental health, reinforcing the therapy's applicability. For example, CBT has been used to assist people with irrational phobias to alter their judgment toward and aversion of fear-instilling stimuli [22], as an antidepressive treatment [23], to treat eating disorders [24,25], and many more.

Accessing CBT

More typically, CBT has been delivered in a clinical environment by a therapist who meets face-to-face with a patient. However, this method can present patients with additional obstacles for various reasons. This could be because of time constraints, prior engagements, or misguided prioritization of well-being. Some patients may suffer with social anxiety, agoraphobia, or physical issues that inhibit them from leaving their home.

There are emergent deliverance methods of CBT being explored, such as telephone CBT [26,27], CBT delivered via texting [28], and Internet-based CBT (iCBT). With the introduction of iCBT, patients became able to communicate and confer with their therapist via a Web-based platform, and treatment can be delivered using Web-based programs.

Although it has been argued that iCBT lacks in efficacy because of the absence of face-to-face interaction with a therapist, Carlbring et al [29] carried out a study that evidenced support for Web-based distribution of treatment for panic disorder because of the subsequent large effect sizes. In this study, face-to-face meetings were substituted with short, once weekly phone calls. The sample consisted of 60 participants who met the criteria for panic disorder as categorized by the *Diagnostic and Statistical Manual of Mental Disorders, 4th edition*. The study measured aspects such as cognitions, physiological sensations experienced by patients with anxiety disorders, degree of agoraphobia, generalized anxiety, level of depression, and quality of life. The study spanned over 10 weeks when follow-up data was collected. Follow-up data was also collected at 9 months post study. Results show that the intervention group improved significantly in all measures between pre- and posttreatment (t_{29}=4.4-8.7, all P<.001). Another study asked participants their opinions of iCBT. Of 105 participants, 76.2% (80/105) reported that they had not particularly felt the absence of face-to-face contact with a therapist, and 94.3% (99/105) would recommend the method to others [30].

Access to iCBT equips patients with the provisos and skills to guide their own therapy. Haarhoff and Kazantzis [31] contested that several components of CBT, such as observing and altering one's own beliefs, feelings, and conduct, can be considered as salient self-help techniques. When completing *homework* that

has been set by a CBT therapist, a patient is essentially contributing to their own well-being using a self-help nature.

Naeem et al [32] contests that self-guided CBT (SG-CBT) can be used as an intervention with minimum direct contact. For patients in need of CBT who face obstacles hindering face-to-face attendance, a self-guided platform to access therapy holds many advantages. Previous studies have shown that SG-CBT can instigate and maintain significant clinical improvements [33]. SG-CBT can enable the avoidance of using mental health services, divert from long waiting lists, bypass obstacles that prevent the use of service, and effectively lower the costs of care for both patients and amenities.

The ubiquity of mobile phones today has become the norm for society. Simultaneously, it is estimated that over 8 billion individuals are connected to the World Wide Web using mobile phones as a platform [34]. Due to the copiousness of mobile phones, the utility of mobile health (mHealth) has never been more endemic. Studies have found that 31% of mobile phone owners use them to access health information; 19% have also installed a mobile app that relates to current medical condition or to manage their health and well-being [35,36].

Some mental health and well-being–focused apps are now integrating mobile technology with CBT by using the basic principles and underpinnings of CBT to enhance the outcome of self-guided therapy. As CBT is an umbrella term for various cognitive and behavioral therapies, the therapeutic content of said apps vary.

Aim

Both the increasing use and pervasiveness of mobile phones and the integration of basic CBT principles into mHealth apps are cause for concern because of the clinically unsupervised nature of the intervention and subsequent outcomes. This review aims to synthesize the extant literature and assess the efficacy and user experience of SG-CBT apps with reference to app content. This review followed the guidelines provided by preferred reporting items for systematic reviews and meta-analyses (PRISMA) and adhered to the published PRISMA checklist [37].

Methods

Search Strategy and Selection of Studies

A comprehensive literature search in relevant bibliographic, Web-based databases was carried out (PsycARTICLES, Google Scholar, Scopus, ProQuest Central, JMIR, and PubMed). Initially, search terms used were not restricted to the title only. They were found within the title, abstract, or full paper. Words searched were “cognitive,” “behaviour,” “behavioural,” “therapy,” “CBT,” “mobile phone(s),” “application,” “randomised,” “controlled,” “trial,” and “RCT.” The conjunction “AND” and the logical operator “OR” were also used in the search terms (Multimedia Appendix 1). The searches were consistently inclusive of “randomised,” “controlled,” “trial” and “RCT.” This ensured that the search did not return studies of other experimental designs. These words were searched because of their direct link with the topical therapy and chosen study design under review.

A vast amount of the search terms were used because of their inclusion in the medical subject headings database. The search terms for this systematic review were stringent because of the focus on one method of therapy, one method of delivery, and one method of experimental design. The reason for this is that the current cultural climate dictates less and less time for individuals to enhance their well-being in a physical setting; therefore, the reliance upon mHealth apps and their instant gratification increases. Moreover, the readily available apps are more beneficial to those suffering from debilitating social phobias or physical ailments. RCTs were the only experimental design reviewed because of the fact that results of the study are reported in a way which sees as many biases removed as possible [38].

From the papers returned by the initial search, abstracts were read to check eligibility. If the abstract of the paper was deemed irrelevant, the paper was retracted from further analyses. If the abstracts were relevant, full text of the paper was then reviewed for eligibility. Papers were discarded if they met the exclusion criteria. Eligible, full text papers had the reference section screened to find further relevant papers. Unpublished studies, dissertations, and gray literature were neither sourced nor included in this review.

The second author carried out the literature search and elected the inclusive papers. The first author and the third author then considered and agreed with the final included studies. The first and the second author carried out the risk of bias assessment.

Inclusion Criteria

The inclusion criterion for the systematic review was inflexible. The studies had to be reported in English, and they ranged from January 2008 to May 2017. This is owing to the fact that apps were only released in 2008 [39]. All studies were required to contain an app that used the basic principles of CBT with an RCT design. As highlighted within the search strategy section, studies reviewed were restricted to those of an RCT design based on their data reporting methods and the comprehensive removal of bias. All studies were reported in peer-reviewed, scholarly journal articles. There were no demographic restrictions. The field of study was restricted to that of mHealth.

Exclusion Criteria

Studies devoid of actual apps were excluded as they were irrelevant to the focus of the review, for example, proposals, reviews, meta-analyses, conferences, and case studies. Dissertations, secondary sources, and gray literature also met the exclusion criteria. This was to ensure that all studies included had an international standard RCT number, essentially meeting the criteria of Consolidated Standards of Reporting Trials [38].

Primary Outcome Measures

This review looked at apps that aim to deliver therapy using the basic principles of CBT and encourage self-guided improvement of one’s well-being. The primary outcome measures were the initial efficacy of the intervention and the overall user experience of the app itself.

Analyses of Effect Sizes

Where the data was attainable, between-group and within-group effect sizes (Cohen *d*) were established using the variance between the pretest and posttest results (within-group effect size) or the variance between the control and intervention group posttest results (between-group effect size) and dividing by the pooled standard deviation. Effect sizes of 0.2 are deemed relatively small. Effect sizes of 0.5 are deemed to be moderate, and those of 0.8 or higher can assume to be associated with large effect sizes [40,41].

Quality Assessment

The overall qualities of included studies were assessed using the Cochrane risk of bias assessment tool [42]. The tool checks studies for biases such as random sequence generation, allocation concealment, incomplete outcome data, selective reporting, and other biases (Multimedia Appendix 2). The blinding criteria of the Cochrane risk of bias assessment tool were not studied as the criteria are almost impossible to abide by when implementing mental health interventions [43]. The Cochrane risk of bias assessment has been highlighted as the most appropriate method of quality assessment for the RCT design studies [44].

Results

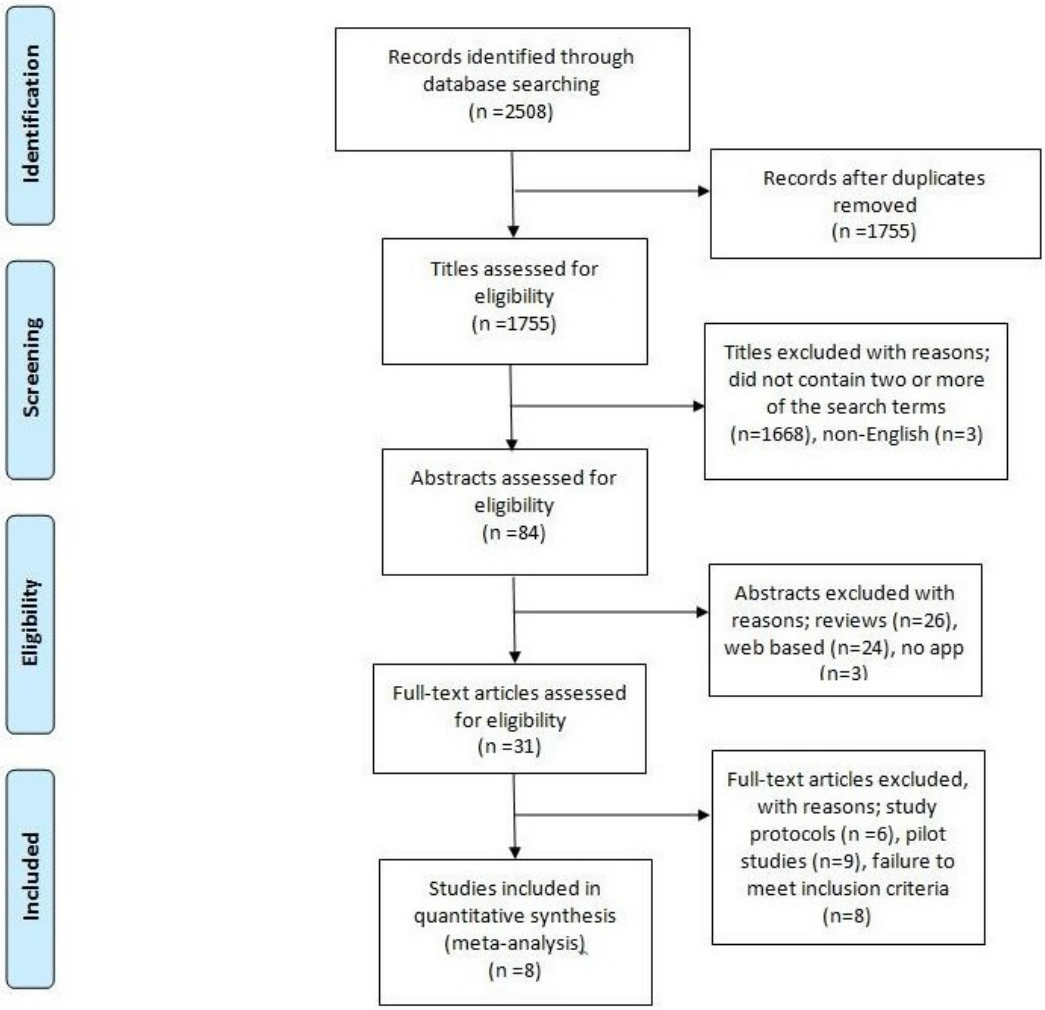
Included Papers

An amassed total of 2508 records were considered through search identification. After removing duplicate records, a total of 1755 records were considered by their title alone. Records were excluded if their title did not contain one or more of the utilized search terms (n=1668). Non-English papers were also excluded at this stage (n=3). Abstracts were then assessed for eligibility, with papers being excluded if they were reviews, Web-based studies, or studies that contained no information of an app being used. The remaining articles (n=31) were deemed eligible for full text review. Following the full text reviews, a further 23 papers were removed because of being pilot studies, study protocols, or failing to meet inclusion criteria. This left a total of 8 studies eligible for review as presented in Figure 1.

Study and Intervention Characteristics

Across all studies there were an overall total of 1794 participants. Each of the included studies utilized an app, 3 of which were pitted against a control group. The remaining 5 studies used the app intervention against another form of treatment or intervention. Some mHealth interventions have remained unnamed within the studies (n=4), whereas others have been coined (n=4).

Figure 1. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram.



All studies (n=8) measured psychological outcomes such as depression (n=4), chronic pain acceptance (n=1), insomnia severity (n=1), stress (n=1), and PTSD symptoms (n=1). Multimedia Appendix 3 provides a comprehensive overview of all inclusive study characteristics.

Risk of Bias

All of the studies employed an RCT design and were overall, deemed to be of low to moderate risk when assessed against Cochrane risk of bias assessment tool. However, there were some studies that had elements of risk.

Some studies failed to report their study's random sequence generation (n=4). Due to this, it was unclear as to whether the appropriate procedure had been adhered to [45-48].

The Cochrane's risk of bias assessment helped researchers to identify further biases. For example, in 2 studies, participants were offered monetary incentives, potentially altering the demand characteristics [47,49]. Another study contained an unequal gender split [45].

The majority of the studies were low risk for the outcome data assessment; however, Whittaker et al's [48] study was deemed to be potentially high risk as only a relatively small fraction of participants read the interventional messages in their entirety. Horsch et al's [50] study had a high attrition rate, and follow-up data collection was taken for the intervention group but not collected for the wait-list control group. Therefore, no comparison could be made for the longer term effects of the intervention. Overall, the studies had high response rates and strong engagement, evidenced in Multimedia Appendix 3.

In this review, all studies employed an RCT design to measure the efficacy of CBT mHealth interventions. Across inclusive studies, the range of the within-group and between-group effect sizes were −0.13 to 1.83 at posttest and 0.03 to 1.44 at follow-up. The largest effects were often seen when comparing the data from pretest to posttest for the app engaged group. These effect sizes diminished at follow-up data collection dependent upon length between posttest and follow-up.

Efficacy

The inclusive studies provide promising results with regard to the integration of the basic principles of CBT in mobile phone apps for mental health. The Sleepcare app, devised by Horsch et al [50], was the first of its kind. The app used CBT for insomnia (CBT-I) via a diary method that aimed to treat the symptoms of insomnia severity. This was to increase access to treatment. A significant interaction of the app was found (P<.01), and there was a large effect on insomnia therapy (d=−0.66). A significant effect upon insomnia severity was also recorded at a 3-month follow-up.

One study carried out by Kuhn et al [49] aimed to meet the needs of the population who suffered with PTSD but were not receiving treatment, by offering them psychoeducational information and evidence-based cognitive behavioral coping tools through an app. Post treatment, participants who used the app as an intervention had significant improvements in PTSD symptoms (P=.04), depression symptoms (P=.005), and psychosocial functioning (P=.007). However, at follow-up tests, it was found that there was no difference between the wait-list participants and the intervention participants. This shows that although the initial results were significant, there was no longevity of efficacy.

Kristjánsdóttir et al [51] found a dearth in the research of chronic pain acceptance and mobile phone self-guided support. Their study explored the efficacy of their unnamed app that was used for 4 weeks and consisted of a diary format and therapist feedback following a chronic pain rehabilitation program. The intervention significantly improved pain acceptance, lowered the catastrophizing of pain, and had a large effect size (d=0.87). At 5 months posttest, the between-group effect sizes were still moderate (d=0.74, P=.003), acceptance of pain (d=0.54, P=.02), and functioning and symptom levels (d=0.75, P=.001). However, it cannot be specified whether the effect sizes were because of the app itself, therapist feedback, or even the rehabilitation program. The effects could be because of the combined participation in each method of therapy.

Ly et al [52] used an unnamed app that produced moderate within-group effects on stress (Perceived Stress Scale [PSS]: d=0.50) and a moderate to large effect between groups (PSS: d=0.62).

Of the 8 studies, half aimed to treat the symptoms of depression (n=4). In a noninferiority trial, Ly et al [45] compared blended treatment, which consisted of a mobile phone and face-to-face therapy, against full behavioral activation treatment for depression. Although the study gave no conclusive findings with regard to the more efficacious treatment, it did find that the blended approach could treat up to twice as many patients because of the ease of access to the app. The MoodHacker app [47] had a significant effect on depression and negative thoughts, and the app MEMO [48] had a large, significant between-group effect on increased positivity when using CBT.

Overall, the collated studies evidenced that CBT delivered via the medium of mobile phones apps, utilized in a self-guided manner, can be efficacious. Half of the studies aimed to treat depression, with generally positive results. mHealth CBT aided the reduction of scaled depressive symptoms and negative thoughts. However, it is salient to note that these results did not maintain significance unless there was continuity of engagement with the app. It was also found that using standard CBT in conjunction with mobile phone deliverance enabled a wider dispersal of the therapy. This is because of the fact that mobile phones are easier to access than health care settings. Both apps that investigated the treatment of PTSD and chronic pain acceptance were innovative and focused upon coping strategies. This suggests that a participant's engagement with the therapy, willingness, and ability to access mHealth platforms is key to the efficacy of CBT delivered using this method.

User Experience

Of the 36 participants that Ly et al [52] included in their study, only 16 adhered to the intervention for the full time span of 6 weeks. In Birney et al's [47] study, participants who received the MoodHacker app intervention used it an average of 16 times (standard deviation [SD] 13.3), which totaled an average of 1.3 hours usage across the study (SD 1.3). A total of 82.4% found

the MEMO app [48] to be useful and stated they would reuse it. A significant proportion of the intervention sample (90.7%) said that they would recommend the app to a friend. The Sleepcare app [50] found that 35 participants filled out >35 diaries, but few participants (n=4) completed <10% of the conversations. Kristjánsdóttir et al [51] found that 83.3% of participants completed 84 entries, yet 23.3% perceived the experience to be a burden.

User experience was measured using varying scales such as time of engagement and personal opinions of a particular app. The results varied. Although participants claimed that they found particular apps useful, this was not reflected by longitudinal engagement. There are multitudinous reasons why participants would deviate from the optimum engagement levels. For example, daily roles and responsibilities such as work, education, or parenting may not allow for optimum use. Other variables such as the lack of Wi-Fi connection and the loss of battery life may have hindered usage. These variables could possibly have skewed results. There were individuals who found diary entries a hindrance rather a help.

Discussion

Limitations

Although comprehensive, this review is not without limitations. This review is selective in its reporting as it only reviews studies employing an RCT design. This is not to refute the designs of excluded studies and their subsequent results, showing the efficacy of mHealth apps. In defense of this, the justification behind the selectiveness is stated within the rationale for review. Additionally, this review analyzed English language papers only; therefore, cross-cultural variations cannot be considered.

With regard to the topical therapy, CBT holds its own limitations. The efficacy of CBT is dependent on many variables. It requires a large amount of commitment, engagement, and cooperation from a patient. It can also be perceived time-consuming because of the regular sessions and frequent homework tasked.

Future Research

Future research ought to work toward the long-term effectiveness of mHealth apps. Longitudinal studies ought to be carried out to assess usage against nonusage for both longer periods of time and greater follow-up periods. The necessity to eradicate technical problems before participant usage is key to fully assess the efficacy of mHealth apps without the interference of study formulated confounding variables.

There is a responsibility for both health professionals and media professionals alike to steer the general public away from the less scientifically created apps and toward regulated and sufficiently formulated and investigated apps. CBT therapists should be consulted and fully involved where apps are created using the principles of CBT to avoid information being misconstrued during usage. The comprehensive creation of mHealth apps with the relevant professionals will create a greater accessibility to effective treatment for the general public. Future apps would benefit from input from multidisciplinary teams during the design and development stages.

From a clinical perspective, medical information from professionals such as doctors, therapists, and counselors is imperative for inclusion in an app. This ensures that the app will distribute updated and medically validated advice and guidance. However, it would also be advantageous to include potential users of the app or those who have engaged with the therapy via varying mediums such as computerized, telephone, and face-to-face. This could provide valuable insights regarding the positive and negative aspects of all methods of deliverance, essentially guiding the development and improving the app. The inclusion of impartial charities and foundations that offer disorder-specific advice and guidance would benefit users. The field will of course take time to develop given the funding implications and time constraints involved with such research projects and the commercial value of other apps.

Conclusions

This review specifically focused upon mHealth apps containing CBT content and RCT designs. In comparison to previous reviews, this review differs because of its focus upon one type of mHealth therapy app. Other reviews have explored mHealth apps in general, with reference to their purpose in health interventions [53]. This review investigated how one specific approach to therapy (CBT) has an impact when delivered as part of an mHealth app.

The results of this review highlight the effectiveness of mHealth apps that use CBT principles. Although the suitability of RCT studies assessing the efficacy of apps has been brought into question for reasons such as the inability to blind a participant to an mHealth intervention, the design has proven effective within this review. It is also worth noting that a patient's motivation and commitment to therapy can determine the outcome of the therapy. Altering maladaptive cognition requires a willingness and devotion that individuals with mental health issues and vulnerability may not always possess. Although standard face-to-face CBT aims to be as personalized as possible, the same principles should be applied during the development of CBT mHealth apps.

Across a range of psychological issues, mHealth apps appear to repeatedly show improvements in symptom severity. Despite the clear effectiveness, the issue of longevity remains. With lower levels of effect at longer time intervals, questions may be raised over the long-term benefits of these mHealth apps. Similarly, the time constraints and funding issues for the creation of said apps is an issue with unregulated apps often taking precedence in the market place. Future research ought to address issues highlighted within this review and work toward more substantial datasets, long-term effects, and professionally created apps with the help of CBT therapists.

Acknowledgments

The authors would like to thank the authors for all data used in the review and the University of Bolton for funding this review through the Ainsworth scholarship.

Authors' Contributions

AR, LC, and JP were involved in the concept and the design of the above review. All authors had major contributions to data analysis and to the write-up and editing of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Searched Terms.

[PDF File (Adobe PDF File), 21KB - jmir_v19i11e399_app1.pdf]

Multimedia Appendix 2

Cochrane's Risk of Bias.

[PDF File (Adobe PDF File), 41KB - jmir_v19i11e399_app2.pdf]

Multimedia Appendix 3

Characteristics of mHealth apps using CBT.

[PDF File (Adobe PDF File), 36KB - jmir_v19i11e399_app3.pdf]

References

1. Hollon SD, Beck AT. Cognitive and cognitive-behavioral therapies. In: Handbook of Psychotherapy and Behavior Change. 4th edition. Oxford, England: John Wiley; 1994:428-466.
2. Olatunji BO, Kauffman BY, Meltzer S, Davis ML, Smits JA, Powers MB. Cognitive-behavioral therapy for hypochondriasis/health anxiety: a meta-analysis of treatment outcome and moderators. Behav Res Ther 2014 Jul;58:65-74. [doi: 10.1016/j.brat.2014.05.002] [Medline: 24954212]
3. Huang FF, Lai YH, Lin CH, Yang HH. Health knowledge insufficiency motivates health information seeking. Int J Web Appl 2016;8(1):1-7 [FREE Full text]
4. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
5. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression. New York: Guilford Press; 1979.
6. Beck AT, Freeman A, Davis DD. Cognitive Therapy of Personality Disorders. 3rd edition. New York: Guilford Press; 2015.
7. Dobson KS. Handbook of Cognitive-Behavioral Therapies. 3rd edition. New York: Guilford Press; 2009.
8. Beck JS. In session with Judith S. Beck: cognitive-behavioral therapy. Cognitive-behavioral therapy. Prim. Psychiatr 2006 Apr;13(4):31-34.
9. Wu Y, Lang Z, Zhang H. Efficacy of cognitive-behavioral therapy in pediatric obsessive-compulsive disorder: a meta-analysis. Med Sci Monit 2016 May 16;22:1646-1653 [FREE Full text] [Medline: 27182928]
10. Abramowitz J, Whiteside SP, Deacon BJ. The effectiveness of treatment for pediatric obsessive-compulsive disorder: a meta-analysis. Behav Ther 2005;36(1):55-63. [doi: 10.1016/S0005-7894(05)80054-1]
11. Sánchez-Meca J, Rosa-Alcázar AI, Iniesta-Sepúlveda M, Rosa-Alcázar A. Differential efficacy of cognitive-behavioral therapy and pharmacological treatments for pediatric obsessive-compulsive disorder: a meta-analysis. J Anxiety Disord 2014 Jan;28(1):31-44. [doi: 10.1016/j.janxdis.2013.10.007] [Medline: 24334214]
12. Watson HJ, Rees CS. Meta-analysis of randomized, controlled treatment trials for pediatric obsessive-compulsive disorder. J Child Psychol Psychiatry 2008 May;49(5):489-498. [doi: 10.1111/j.1469-7610.2007.01875.x] [Medline: 18400058]
13. Romanelli RJ, Wu FM, Gamba R, Mojtabai R, Segal JB. Behavioral therapy and serotonin reuptake inhibitor pharmacotherapy in the treatment of obsessive-compulsive disorder: a systematic review and meta-analysis of head-to-head randomized controlled trials. Depress Anxiety 2014 Aug;31(8):641-652. [doi: 10.1002/da.22232] [Medline: 24390912]

14. Olatunji BO, Davis ML, Powers MB, Smits JAJ. Cognitive-behavioral therapy for obsessive-compulsive disorder: a meta-analysis of treatment outcome and moderators. J Psychiatr Res 2013 Jan;47(1):33-41. [doi: 10.1016/j.jpsychires.2012.08.020] [Medline: 22999486]
15. Cohen JA, Deblinger E, Mannarino AP, Steer RA. A multisite, randomized controlled trial for children with sexual abuse-related PTSD symptoms. J Am Acad Child Adolesc Psychiatry 2004 Apr;43(4):393-402 [FREE Full text] [doi: 10.1097/00004583-200404000-00005] [Medline: 15187799]
16. Schnurr PP, Friedman MJ, Engel CC, Foa EB, Shea MT, Chow BK, et al. Cognitive behavioral therapy for posttraumatic stress disorder in women: a randomized controlled trial. J Am Med Assoc 2007 Feb 28;297(8):820-830. [doi: 10.1001/jama.297.8.820] [Medline: 17327524]
17. Cohen JA, Berliner L, Mannarino A. Trauma focused CBT for children with co-occurring trauma and behavior problems. Child Abuse Negl 2010 Apr;34(4):215-224. [doi: 10.1016/j.chiabu.2009.12.003] [Medline: 20304489]
18. Smith P, Yule W, Perrin S, Tranah T, Dalgleish T, Clark DM. Cognitive-behavioral therapy for PTSD in children and adolescents: a preliminary randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2007 Aug;46(8):1051-1061. [doi: 10.1097/CHI.0b013e318067e288] [Medline: 17667483]
19. Norton PJ, Barrera TL, Mathew AR, Chamberlain LD, Szafranski DD, Reddy R, et al. Effect of transdiagnostic cbt for anxiety disorders on comorbid diagnoses. Depress Anxiety 2013 Feb;30(2):168-173. [doi: 10.1002/da.22018] [Medline: 23212696]
20. Barrett PM, Duffy AL, Dadds MR, Rapee RM. Cognitive-behavioral treatment of anxiety disorders in children: long-term (6-year) follow-up. J Consult Clin Psychol 2001 Feb;69(1):135-141. [Medline: 11302272]
21. DiMauro J, Domingues J, Fernandez G, Tolin DF. Long-term effectiveness of CBT for anxiety disorders in an adult outpatient clinic sample: a follow-up study. Behav Res Ther 2013 Feb;51(2):82-86. [doi: 10.1016/j.brat.2012.10.003] [Medline: 23262115]
22. Paquette V, Lévesque J, Mensour B, Leroux JM, Beaudoin G, Bourgouin P, et al. "Change the mind and you change the brain": effects of cognitive-behavioral therapy on the neural correlates of spider phobia. Neuroimage 2003 Feb;18(2):401-409. [Medline: 12595193]
23. Johnsen TJ, Friborg O. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: a meta-analysis. Psychol Bull 2015 Jul;141(4):747-768. [doi: 10.1037/bul0000015] [Medline: 25961373]
24. Fairburn CG. Cognitive Behavior Therapy and Eating Disorders. New York: Guilford Press; 2008.
25. Agras WS, Fitzsimmons-Craft EE, Wilfley DE. Evolution of cognitive-behavioral therapy for eating disorders. Behav Res Ther 2017 Jan 1;88:26-36. [doi: 10.1016/j.brat.2016.09.004] [Medline: 28110674]
26. Brenes GA, Danhauer SC, Lyles MF, Anderson A, Miller ME. Effects of telephone-delivered cognitive-behavioral therapy and nondirective supportive therapy on sleep, health-related quality of life, and disability. Am J Geriatr Psychiatry 2016 Oct;24(10):846-854 [FREE Full text] [doi: 10.1016/j.jagp.2016.04.002] [Medline: 27421617]
27. Coughtrey AE, Pistrang N. The effectiveness of telephone-delivered psychological therapies for depression and anxiety: a systematic review. J Telemed Telecare 2016 Dec 30:1357633X1668654. [doi: 10.1177/1357633X16686547]
28. Shapiro J, Bauer S. Use of short message service (SMS)-based interventions to enhance low intensity CBT. In: Bennett-Levy J, Richards DA, Farrand P, Christensen H, Griffiths KM, Kavanagh DJ, et al, editors. Oxford Guide to Low Intensity CBT Interventions. USA: Oxford University Press; 2010.
29. Carlbring P, Bohman S, Brunt S, Buhrman M, Westling BE, Ekselius L, et al. Remote treatment of panic disorder: a randomized trial of internet-based cognitive behavior therapy supplemented with telephone calls. Am J Psychiatry 2006 Dec;163(12):2119-2125. [doi: 10.1176/appi.ajp.163.12.2119] [Medline: 17151163]
30. Ruwaard J, Lange A, Broeksteeg J, Renteria-Agirre A, Schrieken B, Dolan CV, et al. Online cognitive-behavioural treatment of bulimic symptoms: a randomized controlled trial. Clin Psychol Psychother 2013;20(4):308-318. [doi: 10.1002/cpp.1767] [Medline: 22298417]
31. Haarhoff BA, Kazantzis N. How to supervise the use of homework in cognitive behavior therapy: the role of trainee therapist beliefs. Cogn Behav Pract 2007;14(3):325-332. [doi: 10.1016/j.cbpra.2006.08.004]
32. Naeem F, Johal R, McKenna C, Rathod S, Ayub M, Lecomte T, et al. Cognitive Behavior Therapy for psychosis based Guided Self-help (CBTp-GSH) delivered by frontline mental health professionals: results of a feasibility study. Schizophr Res 2016 May;173(1-2):69-74. [doi: 10.1016/j.schres.2016.03.003] [Medline: 26971071]
33. Titov N, Dear BF, Johnston L, McEvoy PM, Wootton B, Terides MD, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: a 12-month follow-up of a randomised controlled trial. PLoS One 2014;9(2):e89591 [FREE Full text] [doi: 10.1371/journal.pone.0089591] [Medline: 24586897]
34. Gsmaintelligence. Definitive data and analysis for the mobile industry URL: https://www.gsmaintelligence.com/ [accessed 2017-03-21] [WebCite Cache ID 6p7vztJh2]
35. Fox S, Duggan M. Pewinternet. Washington, DC: Pew Research Center's Internet & American Life Project; 2012. Mobile Health 2012 URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2017-11-12] [WebCite Cache ID 6uuem4Uqv]
36. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]

37. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 2;4:1. [doi: 10.1136/bmj.g7647]
38. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. J Am Med Assoc 1996 Aug 28;276(8):637-639. [Medline: 8773637]
39. Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR Mhealth Uhealth 2015 Mar 18;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
40. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
41. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
42. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [Medline: 22008217]
43. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
44. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med 2015 Feb;8(1):2-10. [doi: 10.1111/jebm.12141] [Medline: 25594108]
45. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 9;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
46. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015 May 26;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
47. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
48. Whittaker R, Merry S, Stasiak K, McDowell H, Doherty I, Shepherd M, et al. MEMO--a mobile phone depression prevention intervention for adolescents: development process and postprogram findings on acceptability from a randomized controlled trial. J Med Internet Res 2012 Jan 24;14(1):e13 [FREE Full text] [doi: 10.2196/jmir.1857] [Medline: 22278284]
49. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
50. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile phone-delivered cognitive behavioral therapy for insomnia: a randomized waitlist controlled trial. J Med Internet Res 2017 Apr 11;19(4):e70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]
51. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013 Jan 7;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
52. Ly KH, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. Internet Interv 2014 Jul;1(3):95-101. [doi: 10.1016/j.invent.2014.06.003]
53. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]

Abbreviations

BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
CBT-I: cognitive behavioral therapy for insomnia
iCBT: Internet-based CBT
mHealth: mobile Health
OCD: obsessive compulsive disorder
PHQ: Patient Health Questionnaire
PRISMA: preferred reporting items for systematic reviews and meta-analyses
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial

SD: standard deviation
SG-CBT: self-guided cognitive behavioral therapy
TF-CBT: trauma-focused cognitive behavioral therapy

Edited by G Eysenbach; submitted 28.07.17; peer-reviewed by M Alyami, S Pezaro, IN Gomez, A Paglialonga; comments to author 24.08.17; revised version received 11.09.17; accepted 11.10.17; published 28.11.17.

Please cite as:
Rathbone AL, Clarry L, Prescott J
Assessing the Efficacy of Mobile Health Apps Using the Basic Principles of Cognitive Behavioral Therapy: Systematic Review
J Med Internet Res 2017;19(11):e399
URL: http://www.jmir.org/2017/11/e399/
doi:10.2196/jmir.8598
PMID:29187342

Viewpoint

Is Connected Health Contributing to a Healthier Population?

Carmen G Loiselle[1,2,3,4*], RN, PhD; Saima Ahmed[3,4*], BSc

[1]Department of Oncology, Faculty of Medicine, McGill University, Montreal, QC, Canada
[2]Ingram School of Nursing, Faculty of Medicine, McGill University, Montreal, QC, Canada
[3]Segal Cancer Centre, Jewish General Hospital, Montreal, QC, Canada
[4]Experimental Medicine, McGill University, Montreal, QC, Canada
[*]all authors contributed equally

Corresponding Author:
Carmen G Loiselle, RN, PhD
Ingram School of Nursing
Faculty of Medicine
McGill University
680 Sherbrooke West #1840
Montreal, QC, H3A 2M7
Canada
Phone: 1 514 398 4163
Email: carmen.loiselle1@mcgill.ca

Abstract

Connected health tools, including mobile phones, incorporate various functions that capture events, direct actions, and make informed decisions based on complex sources of data. Connected health, a term recently proposed by some academics and industry, refers to the development, testing, and integration of smart technology tools into health care. Through these means, connected health creates interconnectivity across various environments, profoundly changing the way we learn, self-regulate, and communicate with one another. In health care, mobile phones enable more precise diagnostics, personalized health recommendations that enhance patient experiences and outcomes while containing health care costs. However, for connected health to achieve its full potential, issues must be addressed pertaining to active engagement in use, privacy, security, and quality, as well as the development of evidence-based guidelines. This commentary discusses these key challenges and explores the promise of connected health, specifically eHealth and mHealth. Anchored within the context of cancer, the authors' area of expertise, the ideas put forward can readily be applied to other health-related disciplines.

(J Med Internet Res 2017;19(11):e386) doi:10.2196/jmir.8309

KEYWORDS

connected health; mHealth; eHealth; uHealth; ubiquitous health

Our world is increasingly interconnected thanks largely to the technology behind tools that link us, gather information, and display data. This profoundly changes the way we learn, connect, and regulate various aspects of the self (eg, physical, psychological, social, and occupational domains). Globally, more than 50% of the world's population owns a mobile device [1]. In the developed world, this number rises to almost 90%, a clear indicator of the significant impact that science and technology have had on our daily lives [1].

Connected health, a newish term recently proposed by some academics and industry, refers to the development, testing, and integration of smart technology tools into health care [2]. When it comes to physical, psychological, or other health-related issues, connected health can drastically change how health information and support are accessed, communicated, monitored, and acted upon. Mobile health (mHealth) apps are being positioned by developers at the very core of convenience and capabilities, allowing with 1-click easy access to our dearest curiosities, needs, worries, and wants. In 2015, there were more than 165,000 mHealth apps available for users to download [3]. Recent reports suggest that these can significantly support patient engagement, self-management, and empowerment [2]. Interestingly enough, however, recent statistics in the United States indicate that merely 12% of health-related apps are responsible for 90% of downloads [3]. One may ask then—why is the wide range of health apps available so underused given significant trends toward fitness and health? What may be important barriers that potential users face? A recent study examining user perceptions found that lack of health app literacy and not knowing about the various options available were the

two most important barriers to adoption [4]—this is even more so for seniors, lower socioeconomic groups, those with less than a high school education, and non-English speakers [2,3]. From a technological point of view, the current myriad of connected health options (eg, health apps) makes it difficult for consumers to choose the most personally relevant and high-quality tools. In the oncology field, the options can be overwhelming to the millions of individuals affected—Belong–Beating Cancer Together, Best Cancer Fighting Foods, Breast Cancer Reference, Cancer Quotes, Cancer Therapy Adviser, ePrognosis, iCancerHealth, Kill The Cancer, MyCancerCoach, My Cancer Manager, and more.

While others point to the paucity of decision aids to inform optimal choice and lack of evidence for significant effects [3], we contend there are more proximal reasons for connected health underuse. The current focus may be overly placed on technology while ignoring the need for a true paradigm shift in health care—one that welcomes, encourages, and requires patients to be activated and engaged in their own wellness and care. In addition, uptake strategies must be in place to ensure that individual engagement in connected health is highest prior to facing health and illness issues. Perhaps a better understanding of the root causes, historically, of the health literacy gap would shed further insight into this important factor. In addition, the prescreening of high-quality apps with hands-on guidance, especially for those less likely to use connected health, is known to increase adoption and maintenance [5]. More engaging features include perceived usefulness, simplicity, and credibility of the sources [4,5]. Therefore, knowledgeable and credible health care providers are in an excellent position as promoters and facilitators of connected health use [6,7,8]. In line with a change in paradigm, there must be a shift in the role of health care providers from fixer to coach, enabling empowerment as much as possible in people facing health challenges, whether acute or chronic.

In terms of quality control, consumers (and many manufacturers) remain unaware of technological standards and regulations to which connected health products must be held accountable [9]. The manner in which app stores are set up, for instance, means that popular apps (but not necessarily the best ones) are more visibly displayed and more readily downloadable. However, lack of supporting evidence means that users often rely solely on subjective 5-star ratings to gauge the quality of apps. Whereas recent frameworks have been put forward to assess connected health quality, such as the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework to evaluate health behavior interventions, the Mobile App Rating Scale (MARS) to access quality, Consolidated Standards of Reporting Trials (CONSORT-EHEALTH) for the standardized reporting of mHealth and electronic health (eHealth) research, and the upcoming Canadian Institutes of Health and Research guidelines [9-15], one cannot underestimate the importance of bringing people to become more acquainted with connected health (ie, increasing the number of users) before devoting too many resources to developing and ranking tool quality—tools that individuals may not even use. With increased adoption over time, consumers will also become savvier in discriminating among the various options available to them.

With regard to eHealth dissemination, the transmission of health-related knowledge over the Internet, misinformation can also be an important issue, with little attention paid to verifying the trustworthiness and timeliness of the information [16,17]. Apple's App Store warns that "medical apps that could provide inaccurate data or information, or that could be used for diagnosing or treating patients may be reviewed with greater scrutiny" [18]. Apps that offer peer-to-peer support with content that is user-generated may be particularly at risk for errors and misinformation. While community and support needs may be readily met, anecdotal accounts are not reliable sources and have been associated with negative effects on health outcomes, health care use, and health care provider-patient relationships [16,19]. Government organizations, associations, and the private sector have begun countering this by creating content on their own that provides relevant, trustworthy, and high-quality information [16].

As we are well aware, health care information is particularly sensitive to privacy and security issues, and health care providers are particularly worried about this [3,20,21]. Whereas most oncologists, for instance, know that patients rely on connected health for cancer information and support, they are uncomfortable recommending particular apps to patients [22]. To this end, the National Institutes of Health Informatics, serving as a link among academia, the health informatics industry, and users, offers a variety of resources as a means of guiding innovation uptake and dissemination.

For specific privacy and security issues, concerned consumers are urged to read privacy statements before using connected health to have more control of their privacy [23]. However, regulations vary widely from country to country and, as of yet, no international standards exist [24]. The US Food and Drug Administration, for example, only regulates "a small subset of mobile medical apps that may impact on the performance or functionality of currently regulated medical devices" [25]. In contrast, the European Commission put forward new legislation as part of their ePrivacy Directive and General Protective Regulation of Personal Data [26]. The Department of Justice will officially enact the regulation on May 25, 2018. It will apply to all computers, mobile phones, and tablets and will require informed consent by users for cookies, a form of Web-based data storage, as well as reporting from Internet and app providers of any personal data breaches [27]. Recently, the addition of blockchain technologies that remove the characteristic of infinite reproducibility of data has drastically reduced worries about data security. Benefiting both institutions and patients alike, they remove the onus to safeguard sensitive health data while facilitating the sharing of data. Developers of patient portals and patient navigation platforms are now relying on such security technology [28].

Ultimately, the large number and varying quality of apps available today—combined with few evidence-based guidelines—undermine connected health uptake. However, universally activating people to own a role in their health, wellness, and care as well as empowering them through ready access to relevant connected technologies will translate into unleashing connected health potential like never before. Through the creation of a relevant repository of personal data, individuals

will be empowered to make more informed health and lifestyle decisions and changes [29].

Editorial Note: JMIR Publications discourages the use of "cHealth" (connected health) as it is not clear whether and how the term differs from eHealth, mHealth or uHealth (ubiquitous health).

References

1. Poushter J. Smartphone ownership and Internet usage continues to climb in emerging economies. Washington: Pew Internet and American Life Project; 2016. URL: http://assets.pewresearch.org/wp-content/uploads/sites/2/2016/02/pew_research_center_global_technology_report_final_february_22__2016.pdf [accessed 2017-10-31] [WebCite Cache ID 6udXqEth3]
2. Hesse B, Greenberg AJ, Rutten L. The role of Internet resources in clinical oncology: promises and challenges. Nat Rev Clin Oncol 2016 Dec;13(12):767-776. [doi: 10.1038/nrclinonc.2016.78] [Medline: 27273045]
3. IMS. Patient adoption of mHealth: use, evidence and remaining barriers to mainstream acceptance.: IMS Institute for Health Informatics; 2015. URL: http://www.imshealth.com/files/web/IMSH%20Institute/Reports/Patient%20Adoption%20of%20mHealth/IIHI_Patient_Adoption_of_mHealth.pdf [accessed 2017-11-03] [WebCite Cache ID 6uhgmqxMC]
4. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
5. 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]
6. Hesse B, Greenberg A, Rutten L. The role of Internet resources in clinical oncology: promises and challenges. Nat Rev Clin Oncol 2016 Dec;13(12):767-776. [doi: 10.1038/nrclinonc.2016.78] [Medline: 27273045]
7. Boulos M, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
8. Loiselle C, Edgar L, Batist G, Lu J, Lauzier S. The impact of a multimedia informational intervention on psychosocial adjustment among individuals with newly diagnosed breast or prostate cancer: a feasibility study. Patient Educ Couns 2010 Jul;80(1):48-55. [doi: 10.1016/j.pec.2009.09.026] [Medline: 19854604]
9. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
10. Glasgow R, Vogt T, Boles S. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [Medline: 10474547]
11. Glasgow RE, Dzewaltowski DA, Estabrooks PA. Assessing the public health impact of health promotion initiatives: RE-AIM evaluation framework. URL: http://re-aim.org/ [accessed 2017-10-31] [WebCite Cache ID 6udZXsB5O]
12. Stoyanov S, Hides L, Kavanagh D. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. J Med Internet Res mHealth eHealth 2015:3.
13. Stoyanov S, Hides L, Kavanagh D, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
14. About eHealth innovations.: Canadian Institutes of Health Research URL: http://www.cihr-irsc.gc.ca/e/47345.html [accessed 2017-10-31] [WebCite Cache ID 6udZzeKX6]
15. 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]
16. Madathil K, Rivera-Rodriguez A, Greenstein J, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
17. Okuhara T, Ishikawa H, Okada M, Kato M, Kiuchi T. Assertions of Japanese websites for and against cancer screening: a text mining analysis. Asian Pac J Cancer Prev 2017 Dec 01;18(4):1069-1075 [FREE Full text] [doi: 10.22034/APJCP.2017.18.4.1069] [Medline: 28547943]
18. App Store review guidelines.: Apple URL: https://developer.apple.com/app-store/review/guidelines/ [accessed 2017-10-31] [WebCite Cache ID 6udaJ5trn]
19. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
20. Schueller S, Washburn J, Price M. Exploring mental health providers' interest in using Web and mobile-based tools in their practices. Internet Interv 2016 May;4(2):145-151 [FREE Full text] [doi: 10.1016/j.invent.2016.06.004] [Medline: 28090438]

21. About NIHI.: National Institute for Health Informatics URL: http://www.nihi.ca/index.php?MenuItemID=23 [accessed 2017-10-31] [WebCite Cache ID 6udaUcTZr]
22. Newnham G, Burns W, Snyder R, Dowling AJ, Ranieri NF, Gray EL, et al. Attitudes of oncology health professionals to information from the Internet and other media. Med J Aust 2005 Aug 15;183(4):197-200. [Medline: 16097920]
23. Charbonneau D. Privacy practices of health social networking sites: implications for privacy and data security in online cancer communities. Comput Inform Nurs 2016 Aug;34(8):355-359. [doi: 10.1097/CIN.0000000000000249] [Medline: 27253081]
24. Boulos M, Brewer A, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
25. Mobile medical applications.: U.S. Food & Drug Administration URL: https://www.fda.gov/medicaldevices/digitalhealth/mobilemedicalapplications/ucm255978.htm [accessed 2017-10-31] [WebCite Cache ID 6udal4zEn]
26. Digital privacy.: European Commission URL: https://ec.europa.eu/digital-single-market/en/online-privacy [accessed 2017-10-31] [WebCite Cache ID 6udanrTXK]
27. EU legislation on cookies.: European Commission URL: http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm [accessed 2017-10-31] [WebCite Cache ID 6udar5yhQ]
28. MyChart. URL: https://www.mychart.ca/ [accessed 2017-11-03] [WebCite Cache ID 6udbDwIBp]
29. Mesko B. Building a healthier lifestyle with the help of digital health.: Media Planet: Future of Health Care URL: http://www.futureofhealthcarenews.com/digital-health/building-a-healthier-lifestyle-with-the-help-of-digital-tech [accessed 2017-10-31] [WebCite Cache ID 6udb310dr]

Abbreviations

RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance
MARS: Mobile App Rating Scale
CONSORT-eHEALTH: Consolidated Standards of Reporting Trials - eHealth extension

Edited by G Eysenbach; submitted 28.06.17; peer-reviewed by P Anderberg, H Melissant, MS Park; comments to author 20.07.17; revised version received 28.07.17; accepted 09.08.17; published 10.11.17.

Please cite as:
Loiselle CG, Ahmed S
Is Connected Health Contributing to a Healthier Population?
J Med Internet Res 2017;19(11):e386
URL: http://www.jmir.org/2017/11/e386/
doi:10.2196/jmir.8309
PMID:29127077

Original Paper

Engagement Within a Mobile Phone–Based Smoking Cessation Intervention for Adolescents and its Association With Participant Characteristics and Outcomes

Raquel Paz Castro[1], M.Sc.; Severin Haug[1], Ph.D.; Andreas Filler[2,3], M.Sc.; Tobias Kowatsch[2], Ph.D.; Michael P Schaub[1], Ph.D.

[1]Swiss Research Institute for Public Health and Addiction, Zurich University, Zurich, Switzerland

[2]Center for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

[3]Energy Efficient Systems Group, University of Bamberg, Bamberg, Germany

Corresponding Author:
Raquel Paz Castro, M.Sc.
Swiss Research Institute for Public Health and Addiction
Zurich University
Konradstrasse 32
Zurich, 8031
Switzerland
Phone: 41 444481180
Fax: 41 444481170
Email: raquel.paz@isgf.uzh.ch

Abstract

Background: Although mobile phone–delivered smoking cessation programs are a promising way to promote smoking cessation among adolescents, little is known about how adolescents might actually use them.

Objective: The aim of this study was to determine adolescents' trajectories of engagement with a mobile phone–delivered smoking cessation program over time and the associations these trajectories have with baseline characteristics and treatment outcomes.

Methods: We performed secondary data analysis on a dataset from a study that compared a mobile phone–delivered integrated smoking cessation and alcohol intervention with a smoking cessation only intervention for adolescents recruited in vocational and upper secondary school classes (N=1418). Throughout the 3-month intervention, participants in both intervention groups received one text message prompt per week that either assessed smoking-related target behaviors or encouraged participation in a quiz or a message contest. Sequence analyses were performed to identify engagement trajectories. Analyses were conducted to identify predictors of engagement trajectory and associations between engagement trajectories and treatment outcomes.

Results: Three engagement trajectories emerged: (1) stable engagement (646/1418, 45.56%), (2) decreasing engagement (501/1418, 35.33%), and (3) stable nonengagement (271/1418, 19.11%). Adolescents who were younger, had no immigrant background, perceived more benefits of quitting smoking, and reported binge drinking preceding the baseline assessment were more likely to exhibit stable engagement. Due to different reach of more engaged and less engaged participants at follow-up, three statistical models (complete-cases, last-observation-carried-forward, and multiple imputation) for the associations of engagement trajectory and smoking outcome were tested. For 7-point smoking abstinence, no association was revealed to be statistically significant over all three models. However, decreasing engagement with the program was associated over all three models, with greater reductions in daily tobacco use than nonengagement.

Conclusions: The majority of tobacco-smoking adolescents engaged extensively with a mobile phone–based smoking cessation program. However, not only stable engagement but also decreasing engagement with a program might be an indicator of behavioral change. Measures to avoid nonengagement among adolescents appear especially necessary for older smokers with an immigrant background who do not drink excessively. In addition, future studies should not only examine the use of specific program components but also users' engagement trajectories to better understand the mechanisms behind behavioral change.

(J Med Internet Res 2017;19(11):e356) doi:10.2196/jmir.7928

KEYWORDS

tobacco; alcohol drinking; adolescent; mobile phones; treatment outcome

Introduction

Tobacco smoking is one of the main contributors to the global burden of disease [1]. A survey of 15- and 16-year-old adolescents covering 36 European countries revealed that 21% considered themselves current smokers [2]. As tobacco use often starts in adolescence, intervening before the development of a substance use disorder gains importance [3].

Mobile phone–based programs for smoking cessation are promising tools for delivering treatment to large numbers of adolescents [3]. Such programs have already been proven more effective than minimal or no intervention in adult smokers [4-8]. Whereas only trends toward the effectiveness of such programs in adolescents have been documented to date [6], studies highlight their acceptance by adolescent smokers with mixed intentions to quit smoking and by adolescent smokers of different genders, educational levels, and immigrant backgrounds [9,10].

Mobile phone–based smoking cessation programs are delivered via apps [11] or texting [10,12], with the greatest difference being the level of engagement demanded by the two approaches. The first demands that users proactively engage with the program, whereas the latter requires users to actively disengage from the program [13]. Engagement, for instance, has been conceptualized in previous studies both as the usage or the subjective experience with the program [14]. For texting-based programs, there is some evidence that the predominant engagers are female [15] and older and that they exhibit lower rates of daily cigarette consumption [16], but none of these studies were conducted on adolescents.

User engagement with different smoking cessation programs has been linked to positive behavioral changes. With Web-based interventions, for instance, higher numbers of visits and page views were associated with abstinence [17-19]. Recent studies on texting-based interventions point in the same direction. In a study by Balmford and Borland [20], the efficacy of a texting-based smoking cessation program was associated with completion of the program. Participants who elected to stop the program were less likely to be abstinent at follow-up. In another study by Heminger et al [15], rather than overall engagement, *postquit* engagement and the use of specific program features such as pledges were specifically predictive of 6-month abstinence. Even more accurately, a study by Christofferson et al [16] identified five different classes of user engagement, which in turn were associated with different levels of interventional success. These investigators extracted two classes of engagement (high engagement and increasing engagement) and three classes of disengagement (rapidly-decreasing engagement, delayed decreasing engagement, and low engagement), demonstrating that participants within the more engaged classes were significantly more likely to be abstinent at weeks 3, 4, and 5 than participants within less engaged classes.

However, there are also studies that question the association between high engagement and positive behavior changes [11,20,21]. For instance, Balmford and Borland [20] found that users with the lowest texting intensity had a greater chance of being abstinent at the 1-month follow-up. The researchers concluded that users tend to be selective as to what they need, which is not to be confused with a lack of motivation. Furthermore, they questioned whether it would be of more help if greater engagement could be achieved among less responsive users. In another pilot study [11] that investigated the use of an app-based smoking cessation program, the total number of actively-viewed quit tips and medication tips was predictive of nonabstinence at 12-week follow-up.

Three methodological issues make the contribution of a user's level of engagement to long-term abstinence somewhat uncertain. First, only one study has reported long-term outcome associations with engagement [15]. Second, setting a quit date and having a quit attempt is an integral component of most smoking cessation programs [11,15,16,18,20]. Such interventions are typically divided into prequit and postquit phases. There is a lack of studies investigating engagement with a mobile phone–based intervention that was matched to stages of change and did not require subjects to set a quit date. This is of special interest, as most adult and adolescent smokers do not report any serious intention to quit within the next month [10,22]. Third, on the other hand, smokers who enroll in such cessation programs already tend to report an intention to quit smoking [15,16,20], which can lead to a self-selection of more engaged and thus, more successful subjects. To our knowledge, no studies have investigated engagement with a mobile phone–based cessation program in proactively-recruited smokers at different stages of change.

Thus, this study aimed to examine trajectories of program engagement associated with long-term outcomes within a randomized controlled trial (RCT) assessing a fully-automated mobile phone–based program for young smokers that was based on the health action process approach (HAPA) stages of change model [23]. In this study, we conceptualized engagement as the usage of the program. We expected to find trajectories of higher and lower program engagement, similar to the study of Christofferson et al [16]. Compared with their study [16], we did not expect a concrete amount of trajectories, as we applied a different analysis method, and our sample was not only constituted by participants intending to quit smoking. Factors that predict engagement and completion of the 3-month program were analyzed to sort out for whom such programs still need to be improved. We hypothesized that being female [15], older age, and smoking at lower daily rates [16] would predict engagement. In addition, this study investigated adolescents' engagement with different features of a mobile phone–based intervention, as identifying more and less influential components of such interventions has recently been raised as a means to help refine other health behavior change programs [24].

Methods

Participants and Procedures

Data for this study were extracted from a two-arm, parallel-group, cluster RCT that used school class as the randomization unit, as detailed elsewhere [9,25]. Students in vocational or upper secondary schools in Switzerland were invited to participate in a technology-based program called *MobileCoach Tobacco (MCT)* if they (1) either smoked on a daily or occasional basis (at least 4 cigarettes in the preceding month and at least one cigarette during the preceding week) and (2) owned a mobile phone . Participating students were reimbursed 10 Swiss francs for participating in the baseline and follow-up assessments and with 0.5 Swiss francs for each of the 11 text message (short service message, SMS) assessments that they answered within the *MCT* program.

In the original trial, the efficacy of an integrated smoking cessation and alcohol intervention (*MCT+*) was tested against a smoking cessation only intervention (*MCT*) for smoking cessation in adolescents. A total of 1471 students from 360 Swiss vocational school classes participated in this study. They were randomly assigned to either the combined program (*MCT+*, n=730) or to the smoking cessation only program (*MCT*, n=741). The original study [9] found no significant difference between the programs in terms of reducing the number of cigarettes used per day (*MCT+* vs *MCT*: −2.7 vs −2.8) or in increasing the 7-day point prevalence of smoking abstinence at follow-up (*MCT+* vs *MCT*: 14.9% vs 14.0%).

The intervention was designed with, and triggered by, the open-source behavioral intervention platform MobileCoach version 1.1 [26]. The original study protocol was approved by the ethics committee of the Faculty of Philosophy at the University of Zurich, Switzerland (date of approval: June 24, 2014). The study was registered at Current Controlled Trials ISRCTN (ISRCTN02427446, assigned September 8, 2014) and executed in full compliance with the Declaration of Helsinki.

Description of MobileCoach Tobacco

The *MCT+* program combined (1) tailored Web-based feedback on individual drinking behaviors delivered directly after completion of the baseline assessment, (2) tailored mobile phone text messages to promote drinking within low-risk limits over a 3-month period, (3) tailored mobile phone text messages to support smoking cessation for 3 months, and (4) the option of receiving twice daily strategies for smoking cessation centered around a self-defined quit date. Only components (3) and (4) of the integrated intervention were delivered to participants in the *MCT* group. The theoretical backgrounds of these intervention components are described elsewhere [9].

The Web-based feedback, intended only for participants in the combined program, was provided immediately after completion of the baseline assessment. It included individually-tailored information (1) about calorie intake based on personal drinking data and (2) age and gender-specific norms on the number of drinks consumed per week, as well as on the individual's frequency of binge drinking.

The alcohol-related text messages provided information on (1) strategies for drinking within low-risk limits and (2) the association between smoking and alcohol consumption. These text messages were sent only to those subjects within the *MCT+* condition who reported binge drinking previous to their baseline assessment, where binge-drinking is equivalent to the consumption of 5 or more drinks on a single occasion for men and 4 or more drinks for women. These text messages were sent on Saturdays at 7 PM on even weeks, whereas on odd weeks they were sent at each particular individual's typical drinking day and time.

The tobacco-related text messages provided information relevant to each subject's individual HAPA stage of change [23]. On the basis of the HAPA stage [23], subjects can be divided into "preintenders" (individuals with no intention to quit smoking) and "intenders" or "actors" (individuals who seriously intend to quit smoking or have already quit). For preintenders, the text messages addressed the benefits of quitting, risks of smoking, and methods for improving self-efficacy. For intenders, the text messages initiated planning processes, whereas for actors they emphasized self-regulatory skills.

During the 3-month *MobileCoach Tobacco* program, participants in both intervention groups received one text message prompt per week that either assessed smoking-related target behaviors or encouraged the subject's participation in a quiz or message contest. These prompts were easily answered by typing a single letter, number, or sentence using the mobile phone's reply function. Every 4 weeks, smoking-related target behaviors, including the person's HAPA stage of change, were assessed through the question "Have you recently smoked cigarettes?," with the following response options: (1) "Yes, and I do not intend to quit" (preintender), (2) "Yes, but I am considering quitting" (preintender), (3) "Yes, but I seriously intend to quit" (intender), or (4) "No, I have already quit smoking" (actor). Furthermore, among preintenders, the number of cigarettes smoked per day or week (depending on smoking status: daily or occasionally) was assessed every 4 weeks. For intenders and actors, the use of strategies to cope with craving, which were individually chosen within the baseline assessment, was assessed: for example, "Did you apply the following strategy recently? When I am at a party, I distract myself from smoking by dancing. Yes (Y) No (N)."

Quizzes were included thrice during the *MCT*, with the questions targeting (1) smoking norms (percentage of smokers within the same age- and gender-specific reference group), (2) the health consequences of smoking cessation (days until positive health consequences after smoking cessation), and (3) expenditures on cigarettes (money spent on cigarettes per year).

A contest that required participants to create a text message to motivate other participants to quit smoking (for nonintenders) or to provide concrete ways to help others quit smoking (for intenders and actors) was conducted twice during the intervention period. The best text message from each of the two categories, rated weekly by a tobacco cessation expert, was sent anonymously to participants in the respective categories after 48 hours.

Finally, additional text messages were offered to subjects who reported having the intention to quit smoking. Intenders and actors were informed biweekly about the option of receiving additional information around a chosen quit date. After entering a scheduled quit date, the program provided up to 2 daily text messages on quit-day preparation and relapse prevention (weeks −1 to +1: 2 daily text messages; weeks +2 and +3: one daily text message).

Measures

Participants took part in a Web-based health survey during a regular class session, by which data on potential predictors of engagement and outcome variables were collected. The sociodemographic characteristics that were assessed were gender, age, educational attainment, and immigrant background. The following common Swiss levels of educational attainment were assessed: (1) none, (2) secondary school, (3) extended secondary school, and (4) technical or high school. We assessed the country of birth of both parents of the students to identify any potential immigrant background. On the basis of this information, participants were assigned to one of the following categories: (1) neither parent born outside Switzerland, (2) one parent born outside Switzerland, or (3) both parents born outside Switzerland. For further analysis, we grouped subjects with either a one- or two-sided immigrant background into a single category for comparison against nonimmigrants.

The health-related variables that were assessed were perceived stress, physical activity, body weight, typical number of drinks consumed per week, and whether any binge drinking had occurred in the month before the baseline assessment. Perceived stress was measured using the following single item: “In the last month, how severely have you felt stressed?” Participants were asked to indicate their response on a 6-point Likert scale that ranged from *not at all* to *very*. Self-reported moderate to vigorous physical activity was measured by a question derived from the Health Behaviour in School Aged Children study [27]: “Outside school, how many hours a week do you exercise or participate in sports that make you sweat or out of breath?” The typical number of drinks consumed weekly was assessed via a 7-day drinking calendar similar to the Daily Drinking Questionnaire [28], for which participants were asked to think about a typical week in the preceding month and record the number of standard drinks they typically consumed each day during that week. Examples of standard drinks containing 12 g to 14 g of ethanol were provided for beer, wine, spirits, alcopops, and cocktails, along with conversion values (eg, three 0.5 L cans of beer=6 standard drinks). Binge drinking was assessed by asking participants to report the number of standard drinks they consumed on their heaviest drinking occasion over the preceding 30 days.

Tobacco smoking status was assessed using the question, “Are you currently smoking cigarettes?” with the following response options: (1) “Yes—I smoke cigarettes daily,” (2) “Yes—I smoke cigarettes occasionally, but not daily,” and (3) “No.” In occasional smokers, we assessed the number of days they typically smoked per month, the total number of cigarettes smoked within the previous 7 days, and the number of cigarettes smoked on a typical smoking day. In daily smokers, we only assessed the mean number of cigarettes smoked per day. For occasional smokers, the average number of cigarettes smoked per day was computed by multiplying the typical number of smoking days per month with the number of cigarettes smoked on a typical smoking day and dividing this product by 30.

Additionally, we assessed the following smoking-related variables: HAPA stage of change and the number of previous quit attempts. Each subject’s HAPA stage of change was assessed using the following question: “Have you recently smoked cigarettes?” with the following response options (1) “Yes, and I do not intend to quit” (preintenders), (2) “Yes, but I am considering quitting” (preintenders), and (3) “Yes, but I seriously intend to quit” (intenders). Subjects were asked about their previous attempts to quit smoking with the question—“Have you ever made a serious attempt to quit smoking?”—for which they were provided the response options “No,” “Yes—once,” and “Yes—more than once.”

Engagement with the program was assessed in terms of the number of program participants who unsubscribed from the program (program attrition), the number of responses to the weekly text message prompts, the percentage of retrieved versus sent media objects within the program, and the number of smokers who entered a quit date and activated the additional quit day preparation program.

Smoking behavior at 6-month follow-up was assessed as the (1) 7-day point prevalence of smoking abstinence and (2) the mean number of cigarettes smoked per day. To assess the 7-day point prevalence of smoking abstinence, subjects were asked to indicate whether they had taken a puff of a cigarette within the 7 days previous to follow-up. The mean number of cigarettes smoked per day was assessed and computed in the same way as at the baseline assessment.

Statistical Analyses

As a first step, we analyzed whether persons who actively unsubscribed from the intervention differed from those who remained in the intervention, applying Pearson chi-square analysis to examine differences in categorical variables and unpaired student *t* tests for continuous variables. Given that the combined intervention was more extensive, we also examined whether program attrition differed as a function of study condition. Participants who had either unsubscribed or did not receive the text messages, as seen from program log files, were excluded from further analysis. We then explored the use of different program features for the total sample and by treatment arm.

Subsequently, we examined engagement trajectories by analyzing answers to weekly prompts, which were identical for both study groups. To this end, we performed sequence analysis using the TraMineR library (version 1.8-8) in R [29]. For each participant, answers to prompts (as described previously) were ordered into a sequence of states (ie, engagement trajectories). Similarities between participants’ state sequences were computed using the optimal matching (OM) distance algorithm. OM is defined as the minimal effort, in terms of insertions, deletions, and substitutions, of transforming one sequence into another. Homogeneous engagement trajectory groups (clusters)

were then constructed from the distance matrix using agglomerative nesting hierarchical clustering and Ward's linkage method. The number of clusters chosen was based on the highest relative loss of inertia (see function HCPC in FactoMineR package [30]) and upon the quality of the clusters according to the average silhouette width (ASW) [31]. The ASW ranges from −1 to +1 and can be interpreted as the degree of coherence among assignments to clusters: a high degree of coherence (close to 1) indicates large between-group distances and strong within-group homogeneity.

Upon detecting different engagement trajectories, we examined for baseline differences between the clusters. Subsequently, we conducted multinomial logistic regression analysis to identify predictors of clusters characterized by lower engagement trajectories, compared with those with higher engagement trajectories. Initially, separate univariate multinomial logistic regression analyses were performed (subsequently referred to as univariate analyses) to evaluate potential predictors of engagement trajectories. After these univariate analyses, multivariate prediction models were developed. As suggested by Hosmer et al [32], variable selection consisted of the following steps: (1) significant predictors (P<.05) identified during univariate analyses were entered into the preliminary multivariate model; (2) variables that were nonsignificant at P>.05 were removed, one at a time, starting with those with the highest P values (backward selection); and (3) to account for suppressor effects, the resulting model was verified by adding the aforementioned excluded variables, separately, to the regression model. Only variables that were significant at P<.05 were retained in the final multinomial regression model (forward selection).

Finally, we compared smoking outcomes between participants grouped by their engagement trajectory. As participants were nested in school classes, we conducted a generalized linear mixed model for the 7-day point prevalence of abstinence. For changes in consumed cigarettes per day, we conducted a linear mixed model. Engagement trajectory was included as an independent variable (fixed effect) and school class as a single random effect (random intercept). These analyses were conducted using the lme4 library (version 3.2.1) in R [33] on the following three statistical models because of the disparate reach of more engaged than less engaged participants at follow-up: (1) a complete case analysis (CCA) dataset, (2) a last-observation-carried-forward (LOCF) dataset, and (3) an intention-to-treat (ITT) dataset. Details of outcome analysis and missing data imputation procedures are provided in Haug et al [9]. R version 3.3.3 (The R Foundation for Statistical Computing) was used to perform all sequence analyses and outcome analyses, whereas the Statistical Package for the Social Sciences (SPSS) version 22 (IBM Corp) was used for all other analyses. All statistical tests were two-tailed, with P<.05 set as the criterion for statistical significance.

Results

Participants

Figure 1 depicts the progression of participants through the trial. Of the original 1471 study participants, 1418 (96.39%) completed the program. Those who failed to complete their intervention had either signed off (31/1471, 2.11%) or discontinued the intervention because of technical problems (22/1471, 1.49%). No significant baseline differences were observed between those who did and did not complete the intervention. Program attrition also did not differ between the two treatment arms, with 13 of the 741 (1.8%) participants in the *MCT* choosing to unsubscribe compared with 18 of the 730 (2.5%) assigned to the *MCT+* (χ^2_2=0.9, P=.34). Of the 1418 participants analyzed for this study, 863 (60.86%) were female. The reported mean age was 18.6 (standard deviation [SD] 3.1). More than half (740/1418, 52.18%) reported either a one-sided or two-sided immigrant background, and almost all (1180/1418, 83.21%) had reached at least the lowest educational degree (ie, secondary school). Two-thirds of the sample (1083/1418, 76.37%) took part on the follow-up assessment; 538 of the 712 (75.6%) participants were assigned to the *MCT* and 545 of the 706 (77.2%) were assigned to the *MCT+*.

Use of Different Program Features

Table 1 summarizes different program use characteristics across the total sample and by intervention group. Participants answered a mean of 6.6 (SD 3.5) out of 11 text message prompts. Each participant received between 3 and 5 text messages containing media objects (videos and pictures) that had to be downloaded. On average, participants downloaded 20.5% (SD 31.5) of the received media content. Participants in the *MCT+* downloaded media content significantly more often than their *MCT* counterparts (23.6% vs 17.9%, P<.001). Roughly half of the subjects answered all or almost all of their text message prompts. The fewest answers were recorded for the contest prompt at week 8 (24.9%) and for the HAPA stage query at week 10 (42.9%).

Engagement Analysis

Our inspection of answer behavior over the 3-month intervention revealed different types of engagement trajectory. Some participants exhibited a stable answer pattern (either usually answered or almost never answered text messages). Other participants displayed irregular trajectories. The highest relative loss of inertia measure suggested the following three-cluster solution: cluster 1=stable engagement (SE), cluster 2=decreasing engagement (DE), and cluster 3=stable nonengagement (SNE). On the basis of the ASW, the quality of the three clusters ranged from poor (cluster 2=−0.02) to good (cluster 1=0.55) and excellent (cluster 3=0.70). The low ASW for cluster 2 was because the engagement trajectories included within the cluster differed to a great extent. Some subjects answered text messages only in the beginning, whereas others answered depending on the topic. There were also some participants who only started to answer text messages at the end of the program (Figure 2). As the common element within all these trajectories included in cluster 2 is their instability, the three-cluster solution was considered adequate for the purposes of this study.

Table 1. Use of program components by the overall study sample and by study group.

Program components	All (N=1418)	*MCT* (n=712)	*MCT*+ (n=706)	*P* value
Questions answered, mean (SD[a])[b,c]	6.6 (3.5)	6.5 (3.6)	6.8 (3.5)	.22
Percentage of media objects viewed or of media objects sent, mean (SD)[b,c]	20.8 (31.5)	17.9 (32.6)	23.6 (30.0)	<.001
Answer to quizzes[d,e], n (%)				
Quiz costs (week 1)	975 (68.76)	481 (67.6)	494 (70.0)	.33
Quiz health (week 5)	898 (63.32)	438 (61.5)	460 (65.2)	.16
Quiz norms (week 9)	863 (60.86)	429 (60.3)	434 (61.5)	.64
Answer to HAPA[f] stage of change[d,e], n (%)				
Stage 1 (week 2)	1206 (85.04)	598 (83.9)	608 (86.1)	.26
Stage 2 (week 6)	975 (68.76)	485 (68.7)	490 (68.8)	.96
Stage 3 (week 10)	609 (42.94)	295 (41.4)	314 (44.5)	.25
Answer to smoking-related questions[d,e], n (%)				
CPD[g] or CPW[h] or coping strategy (week 3)	992 (69.95)	490 (68.8)	502 (71.1)	.35
CPD or CPW or coping strategy (week 7)	876 (61.77)	444 (62.4)	432 (61.2)	.65
CPD or CPW or coping strategy (week 11)	749 (52.82)	361 (50.7)	388 (55.0)	.11
Answer to contest, n (%)[d,e]				
Motivational or quit contest (week 4)	626 (44.14)	303 (42.6)	323 (45.8)	.23
Motivational or quit contest (week 8)	353 (24.89)	175 (24.6)	178 (25.2)	.78
Setting of a quit date[d,e]	156/475 (32.8)	79/239 (33.1)	77/236 (32.6)	.92

[a]SD: standard deviation.

[b]*t* test.

[c]Degrees of freedom=1416.

[d]Chi-square test.

[e]Degrees of freedom=1.

[f]HAPA: health action process approach.

[g]CPD: cigarettes smoked per day.

[h]CPW: cigarettes smoked per week.

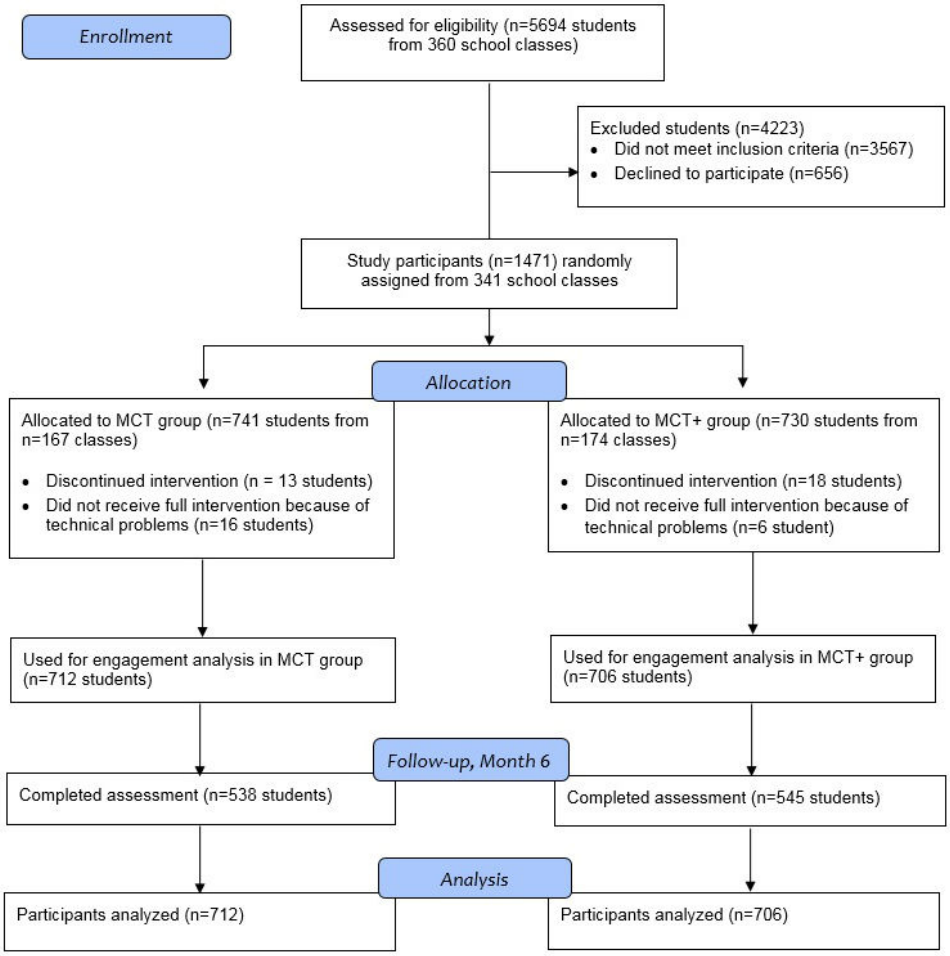
Figures 2 and 3 describe the three clusters in different ways. The first figure displays the response or nonresponse of individuals to each of the 11 prompts within the different clusters. Figure 2 highlights the prototype engagement trajectory within each of the three clusters. The typical participant within cluster 1 (SE) answered to almost all text messages, except for the second request to send their own message to motivate other participants to quit smoking or remain cigarette free. The typical participant within cluster 3 (SNE) did not respond to any of the prompts. Meanwhile, the typical participant within cluster 2 (DE) did not reply to the two message contests and exhibited a steadily decreasing response rate. This last pattern is associated with the repetition of questions, such as queries relating to the person's HAPA stage of change and cigarettes per day or per week.

Predictors of Engagement Trajectory

Table 2 summarizes the baseline characteristics of participants by engagement trajectory. There were significant differences between the three clusters with regard to age (*P*=.006), immigrant background (*P*<.001), educational attainment (*P*=.04), binge drinking (*P*<.001), HAPA stage of change (*P*<.05), and self-perceived benefits of quitting (*P*<.001).

Table 3 shows which of the aforementioned variables were predictive of engagement trajectory within the multivariate model. Being older (Odds ratio [OR]=1.05, *P*=.04) and having an immigrant background (OR=0.76, *P*=.02) predicted a decreasing engagement with the program compared with a stable engagement. Furthermore, participants who perceived more benefits of quitting were more likely to display stable than decreasing engagement with the program (OR=0.52, *P*=.007). Compared with stable engagement, nonengagement was predicted by immigrant background (OR=0.47, *P*<.001) and binge-drinking behavior (OR=1.54, *P*=.005). Being a stable nonengager was more likely than being a stable engager, when participants reported an immigrant background and no binge drinking within the month previous to baseline.

Figure 1. Participants' progress through the trial.



Engagement Trajectories and Smoking Behavior

Treatment outcome by type of engagement trajectory and comparisons of outcomes between engagement trajectories are summarized in Table 4. Reach at follow-up differed importantly between stable engagers (84.3%, 545/646), decreasing engagers (74.5%, 373/501), and nonengagers (59.0%, 160/271). Due to this, three statistical models were tested. Only the reduction in cigarettes per day among decreasing engagers differed significantly from stable nonengagers under the CCA, LOCF, and ITT assumptions (CCA: beta=.65, *P*=.02; LOCF: beta=.43, *P*=.01; and ITT: beta=.54, *P*=.03). Decreasing engagers smoked significantly fewer cigarettes per day at the end of the intervention than nonengagers.

With respect to the 7-day point prevalence of abstinence at 6-month follow-up, no comparison revealed a significant difference under all three assumptions. On ITT analysis, the odds of being abstinent at follow-up was higher among nonengagers than engagers (OR=1.32, *P*=.02). But this finding must be interpreted with caution, as bias in the multiple imputation of missing data seems probable because of the different amount of available information at follow-up. Caution is also warranted, as under the missing-as-smoker assumption, the odds for being abstinent turn in the opposite direction.

Figure 2. The prototype engagement trajectory within each cluster. Columns represent the 11 prompts that could be answered by the participants. Black boxes represent nonreplies, and gray boxes represent replies.

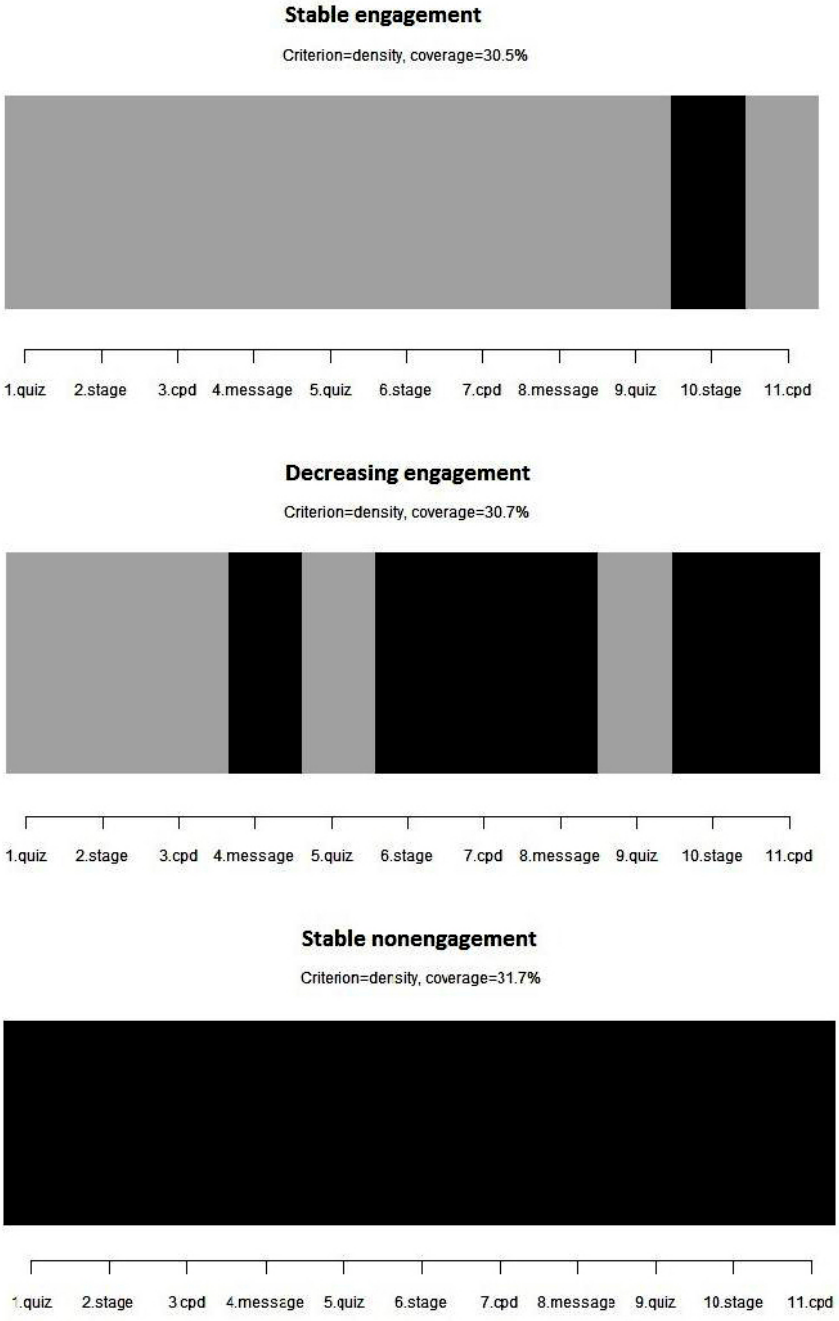


Figure 3. Individual engagement trajectories within each of the three clusters. Rows represent participants and columns represent the 11 prompts that could be answered by the participants. Black boxes represent nonreplies, and gray boxes represent replies.

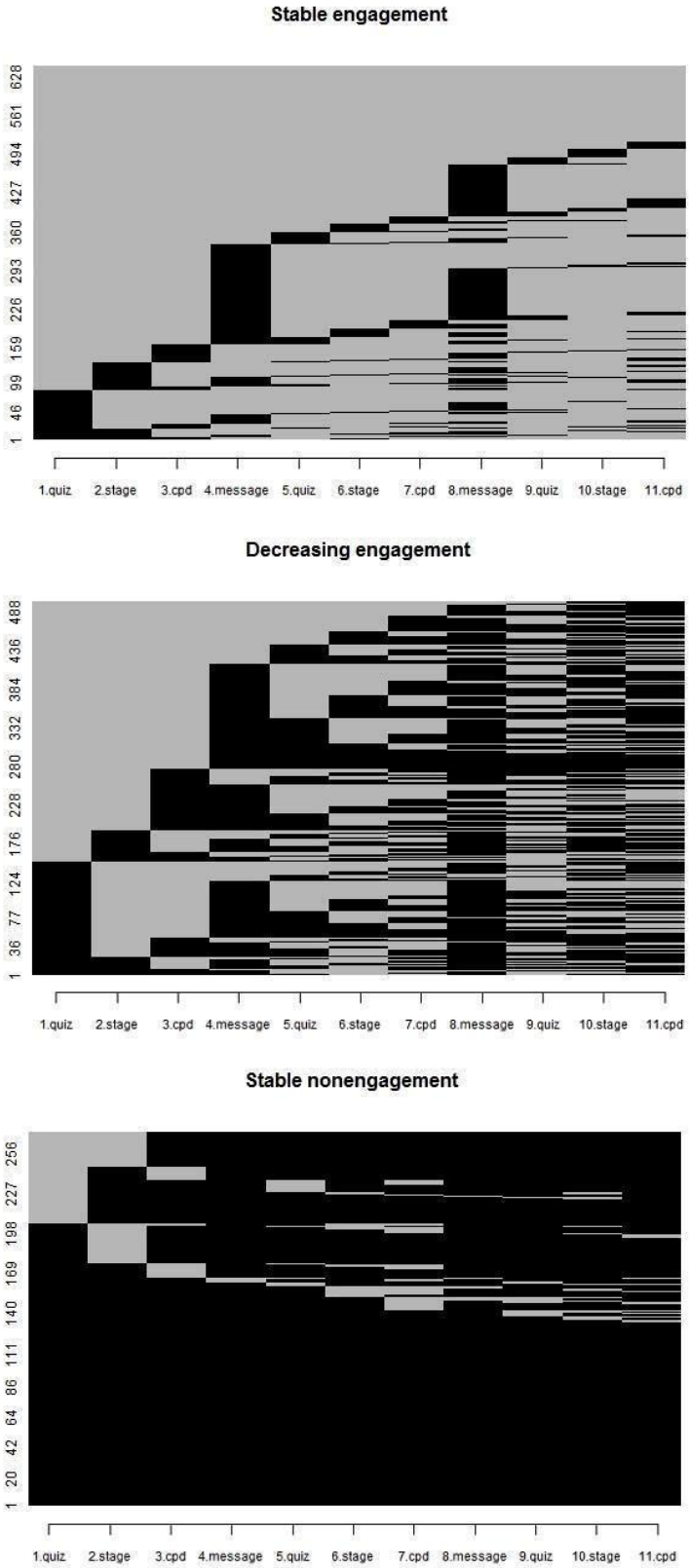


Table 2. Demographic and health behavior characteristics of the study sample by engagement trajectory.

Demographic characteristics	All (N=1418)	Stable engagement (n=646)	Decreasing engagement (n=501)	Stable nonengagement (n=271)	*P* value
Intervention group[a,c], n (%)					
MCT[g]	712 (50.21)	319 (49.4)	248 (49.5)	145 (53.5)	.48
MCT+[h]	706 (49.78)	327 (50.6)	253 (50.5)	126 (46.5)	
Female sex[a,c], n (%)	863 (60.86)	398 (61.6)	293 (58.5)	172 (63.5)	.35
Immigrant background[a,c], n (%)	740 (52.18)	293 (45.4)	268 (53.5)	179 (66.1)	<.001
Age in years, mean (standard deviation [SD[f]])[b,c]	18.6 (3.1)	18.4 (2.8)	18.8 (3.4)	18.9 (2.9)	.006
Educational level[a,e], n (%)					
Secondary school	1180 (83.21)	555 (85.9)	401 (80.0)	224 (82.7)	.04
Vocational school	189 (13.32)	67 (10.4)	86 (17.2)	36 (13.3)	
Technical or high school or university	30 (2.11)	13 (2.0)	10 (2.0)	7 (2.6)	
Unknown	19 (1.33)	11 (1.7)	4 (0.8)	4 (1.5)	
Hours of moderate to vigorous extracurricular physical activity per week, mean (SD)[b,c]	3.5 (3.6)	3.4 (3.4)	3.7 (3.8)	3.2 (3.6)	.08
Number of alcoholic drinks consumed per week, mean (SD)[b,c]	9.9 (12.1)	10.2 (12.1)	10.3 (11.8)	8.4 (12.5)	.09
Binge drinking[a,c], n (%)					
No	465 (32.79)	191 (29.6)	158 (31.6)	116 (42.8)	<.001
Yes	952 (67.13)	455 (70.4)	342 (68.4)	155 (57.2)	
Tobacco smoking status[b], n (%)					
Daily smoker	1075 (75.81)	476 (73.7)	390 (77.8)	209 (77.1)	.22
Occasional smoker	343 (24.18)	170 (26.3)	111 (22.2)	62 (22.9)	
Number of cigarettes smoked per day, mean (SD)[b,c]	10.1 (7.3)	9.9 (7.3)	10.5 (7.4)	10.0 (7.1)	.32
Stage of change[a,d], n (%)					
No intention to quit	396 (27.92)	200 (31.0)	124 (24.8)	72 (26.8)	.03
Considering quitting	825 (58.18)	372 (57.6)	291 (58.2%)	162 (60.2)	
Serious intention to quit	194 (13.68)	74 (11.5)	85 (17.0)	35 (13.0)	
Benefits of quitting smoking, mean (SD)[b,c]	1.38 (0.3)	1.41 (0.3)	1.36 (0.3)	1.37 (0.3)	<.001
Previous quit attempts[a,d], n (%)					
None	507 (35.75)	247 (38.2)	173 (34.6)	87 (32.5)	.10
One	608 (42.87)	276 (42.7)	205 (41.0)	127 (47.4)	
Two or more	299 (21.08)	123 (19.0)	122 (24.4)	54 (20.1)	

[a]χ^2 test.

[b]*F* value.

[c]Degrees of freedom=2.

[d]Degrees of freedom=4.

[e]Degrees of freedom=6.

[f]SD: standard deviation.

[g]MCT: smoking cessation only program.

[h]MCT+: integrated smoking cessation and alcohol intervention.

Table 3. Predictors of engagement trajectory. R^2=.04 (Cox and Snell) and R^2=.05 (Nagelkerke). Model χ^2_8=59.8, P<.001.

Predictors of engagement trajectory	Beta (SE[a])	*P* value	OR[b] (95% CI)
Stable engagement (ref[c]) versus decreasing engagement			
Intercept	−.05 (.57)	.93	
Age in years	.05 (.02)	.04	1.05 (1.003-1.09)
Immigration background (ref yes)	−.28 (.12)	.02	0.76 (0.59-0.96)
Binge drinking (ref yes)	.01 (.13)	.98	1.00 (0.78-1.30)
Benefits of quitting smoking	−.66 (.24)	.007	0.52 (0.32-0.84)
Stable engagement (ref) versus stable nonengagement			
Intercept	−1.07 (.68)	.12	
Age in years	.05 (.03)	.05	1.05 (0.99-1.10)
Immigration background (ref yes)	−.76 (.15)	<.001	0.47 (0.35-0.63)
Binge drinking (ref yes)	.43 (.15)	.005	1.54 (1.14-2.08)
Benefits of quitting smoking	−.37 (.29)	.21	0.70 (0.39-1.24)

[a]SE: standard error.

[b]OR: odds ratio.

[c]ref=reference category.

Table 4. Comparison of treatment outcomes between different engagement trajectories. Descriptive data are based on complete cases. Test value for continuous outcome= *t*-value; for dichotomous outcome=z value.

Engagement trajectory	Mean (SD)	Difference in cigarettes per day[a]			n (%)	7-day point prevalence of smoking abstinence[b]		
		CCA[c] (*P* value)	LOCF[d] (*P* value)	ITT[e] (*P* value)		CCA (*P* value)	MAS[f] (*P* value)	ITT (*P* value)
SE[g] (ref[h])	2.36 (5.5)	.25 (.34)	−.19 (.24)	.07 (.77)	73 (13.4)	1.18 (.16)	0.96 (.76)	1.32 (.02)
SNE[i]	2.43 (5.6)				29 (18.1)			
SE (ref)	2.36 (5.5)	.52 (.01)	.14 (.32)	.40 (.04)	73 (13.4)	1.05 (.58)	0.97 (.77)	1.11 (.21)
DE[j]	3.44 (7.1)				54 (14.5)			
SNE (ref)	2.43 (5.6)	.65 (.02)	.43 (.01)	.54 (.03)	29 (18.1)	0.91 (.46)	0.99 (.97)	0.87 (.25)
DE	3.44 (7.1)				54 (14.5)			
SE (ref)	2.36 (5.5)	.25 (.046)	.00 (.99)	.17 (.18)	73 (13.4)	1.06 (.31)	0.98 (.74)	1.09 (.08)
SNE and DE	2.94 (6.4)				83 (15.6)			

[a]beta

[b]odd ratio.

[c]CCA: complete-case dataset.

[d]LOCF: last-information-carried-forward.

[e]ITT: intention-to-treat dataset.

[f]MAS: missing-as-smoker.

[g]SE: stable engagement.

[h]ref.: reference category.

[i]SNE: stable nonengagement.

[j]DE: decreasing engagement.

Discussion

Principal Findings

Using a proactively-recruited sample of smoking adolescents with mixed intentions to quit smoking, this study examined (1) the use of different components of a mobile phone–based smoking cessation program, (2) different prototypes of engagement trajectory, (3) the association between engagement trajectories and adolescent characteristics, and (4) the association between engagement trajectories and treatment outcomes.

The main findings are as follows: (1) the components of the mobile phone–based smoking cessation program were used over the 3-month intervention in a regular way, with quizzes being the component with the highest participation rate and repeated smoking-related assessments the least-used component, (2) three distinct engagement trajectories emerged: two characterized by higher levels of engagement, stable and decreasing engagement, and one by a lower level of engagement: stable nonengagement, (3) adolescents who were younger, had no immigrant background, perceived more benefits of quitting smoking, and reported binge drinking preceding their baseline assessment were more likely to exhibit a stable engagement trajectory throughout the intervention, and (4) subjects who displayed a decreasing engagement pattern generally reduced their daily tobacco use more than subjects whose level of engagement was low.

This is the first study to examine engagement with a mobile phone–based smoking cessation intervention among adolescents. As expected, trajectories of higher and lower engagement were identified. We found similar results among adolescents as for adults [16,18,20]. We also identified a cluster of people who fully committed to the program, as in the study by Balmford and Borland [20]. Similarly, this study replicates three of the five engagement clusters detected by Christofferson et al [16]. Whereas our cluster-solution was less fine-grained, the clusters were significantly different with respect to baseline characteristics, contrary to those reported by Christoffersons et al [16]. Distinct groups are essential if interventions have to be adapted to different types of engager.

Furthermore, this study was the first to examine factors that predict stable engagement with a mobile phone–based smoking cessation program among adolescents. Other than expected from previous studies on mobile phone–based programs for adult smokers [15,16], engagement was not related to gender. This could be explained by the gender-specific tailoring which was undertaken for *MobileCoach Tobacco* (eg, the feedback on gender-specific drinking norms). Interestingly, in adolescents, being younger was associated with higher levels of engagement versus being older among adults [16]. This result suggests a quadratic relationship between age and engagement. Younger and older people might become more engaged for a variety of reasons that include the program being more novel to them, having more free time, or being more likely to commit to tasks in general. Contrary to our assumptions based on studies in adults, lower rates of daily cigarette consumption were not associated with higher engagement.

This study revealed three further factors, besides age, to be predictive of engagement among adolescents: the individual's immigrant background, their personal outcome expectancies with respect to quitting smoking (ie, the benefits of quitting), and whether or not they previously engaged in binge drinking. An association between immigrant background and use of the program also was identified in a study by Businelle et al [11] that investigated the feasibility and effectiveness of an app-based smoking cessation intervention among socioeconomically disadvantaged adults. Especially non-white participants used the two information-delivering features of the app, which were tips and information about medication for quitting. Future studies should investigate whether tailoring mobile phone–based interventions to a person's immigrant background impacts the intervention's effectiveness. In particular, it has to be examined whether immigrants show less engagement with mobile phone–based programs because of poorer lexical-grammatical skills [34] or because of different interests and socialization than nonimmigrants.

Compared with previous research, the current findings underline the relevance of hazardous alcohol use in predicting engagement with a smoking cessation program. Recent studies on tobacco interventions [35-37] have already highlighted the underestimated role of combined alcohol and tobacco use among adolescents and its association with intervention effectiveness. Not only might mobile phone–based smoking cessation programs be more effective in adolescents who smoke and binge drink [9,35-37], they also could be more attractive to those adolescents. As such, measures are needed to make smoking cessation programs more attractive for adolescents who smoke but do not drink excessively.

Contrary to previous work on adult smokers [15,16,18], we were not able to certainly discern whether more engaged subjects were more likely to be cigarette abstinent after the intervention. This was because of the different reach at follow-up of more engaged compared with less engaged participants. The only stable finding over all statistical assumptions was that a decreasing engagement trajectory was associated with a greater reduction in daily tobacco use than a stable nonengagement trajectory. This result suggests that not only stable engagement but also decreasing engagement might be an indicator of behavior change. As illustrated by other studies [15,20], disengaging from an intervention might not necessarily mean disengaging from behavioral change. Instead, it could indicate a shift from extrinsic to intrinsic motivation [38]. These results support the claim by Yardley et al [39] to examine ways of improving "effective engagement" in subjects rather than simply more engagement, with "effective engagement" defined as sufficient engagement with the treatment to achieve intended outcomes.

One major challenge of future research, however, will be to sort out which kind of intervention is apt for nonengagers. One starting point will be to adapt smoking cessation programs for adolescents to address immigrant backgrounds and drinking behaviors to prevent stable nonengagement and thereby, potentially enhance treatment effectiveness. Considering that most stable nonengagers were more highly motivated to quit smoking at baseline than most stable engagers, one question to

answer will be whether actions must be undertaken to increase active participation or not.

Limitations

The findings of this study must be interpreted in view of its limitations. First, only answers to weekly prompts were included for engagement analysis. Other components of the program—such as downloading media content, setting a quit date, and extracurricular texting behaviors—were not included in our analyses; these components could all be analyzed to determine their own predictive values, similar to [15,40]. However, our rationale for selecting answers to prompts that were identical for all participants was to render our intergroup comparisons more interpretable. Second, that answers to weekly prompts were rewarded with 0.50 Swiss francs to cover the expenses of the adolescents might have influenced the adolescents' likelihood of responding. Third, as already emphasized by Heminger et al [15], quantity and quality of answers to prompts could qualitatively differ (eg, a smoker who replies to all smoking-related prompts and indicates greater daily use of cigarettes). Rather than just analyzing registered events, future qualitative work should investigate whether the content of answers is associated with treatment outcomes. In addition, qualitative research should further investigate the different forms of motivation underlying engagement trajectories among smokers. As stated elsewhere [11], some highly engaged participants might have seen the program as integral to maintaining abstinence, whereas other nonabstinent smokers may have remained highly engaged to prepare for future smoking cessation attempts or merely to receive the offered remuneration. Finally, this study relied on self-report data of smoking behavior, which bears the risk that the results may have been influenced by social desirability.

Conclusions

In summary, in our study, adolescents who smoked engaged to a large extent with a mobile phone–based smoking cessation program, irrespective of their initial intention to quit smoking. Decreasing engagement was in turn clearly associated with better long-term treatment outcomes. Further efforts should be undertaken to increase program engagement among older smokers, with an immigrant background, who do not drink excessively. In addition, future studies should not only examine the use of specific program components but also users' engagement trajectories to better understand the mechanisms behind behavioral change.

Conflicts of Interest

None declared.

References

1. Feigin VL, Roth GA, Naghavi M, Parmar P, Krishnamurthi R, Chugh S, Global Burden of Diseases, Injuries and Risk Factors Study 2013 and Stroke Experts Writing Group. Global burden of stroke and risk factors in 188 countries, during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet Neurol 2016 Aug;15(9):913-924. [doi: 10.1016/S1474-4422(16)30073-4] [Medline: 27291521]
2. Espad. 2016. ESPAD report 2015. Results from the European school survey project on alcohol and other drugs URL: http://www.espad.org/sites/espad.org/files/ESPAD_report_2015.pdf [accessed 2017-10-02] [WebCite Cache ID 6tuaZQHc3]
3. Marsch LA, Borodovsky JT. Technology-based interventions for preventing and treating substance use among youth. Child Adolesc Psychiatr Clin N Am 2016 Oct;25(4):755-768. [doi: 10.1016/j.chc.2016.06.005] [Medline: 27613350]
4. Scott-Sheldon LA, Terry DL, Carey KB, Garey L, Carey MP. Efficacy of expectancy challenge interventions to reduce college student drinking: a meta-analytic review. Psychol Addict Behav 2012 Sep;26(3):393-405 [FREE Full text] [doi: 10.1037/a0027565] [Medline: 22428862]
5. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
6. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: a meta-analysis. Prev Sci 2015 Feb;16(2):181-188. [doi: 10.1007/s11121-014-0498-7] [Medline: 24930386]
7. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
8. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
9. Haug S, Paz Castro R, Kowatsch T, Filler A, Schaub MP. Efficacy of a technology-based, integrated smoking cessation and alcohol intervention for smoking cessation in adolescents: results of a cluster-randomised controlled trial. J Subst Abuse Treat 2017 Nov;82:55-66. [doi: 10.1016/j.jsat.2017.09.008]
10. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013 Aug 16;15(8):e171 [FREE Full text] [doi: 10.2196/jmir.2636] [Medline: 23956024]
11. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. J Med Internet Res 2016 Dec 12;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]

12. Borland R, Balmford J, Benda P. Population-level effects of automated smoking cessation help programs: a randomized controlled trial. Addiction 2013 Mar;108(3):618-628. [doi: 10.1111/j.1360-0443.2012.04091.x] [Medline: 22994457]
13. Suffoletto B. Text message behavioral interventions: from here to where? Curr Opin Psychol 2016 Feb 1;9:16-21. [doi: 10.1016/j.copsyc.2015.09.012] [Medline: 26665157]
14. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
15. Heminger CL, Boal AL, Zumer M, Abroms LC. Text2Quit: an analysis of participant engagement in the mobile smoking cessation program. Am J Drug Alcohol Abuse 2016 Jul;42(4):450-458. [doi: 10.3109/00952990.2016.1149591] [Medline: 27120396]
16. Christofferson DE, Hertzberg JS, Beckham JC, Dennis PA, Hamlett-Berry K. Engagement and abstinence among users of a smoking cessation text message program for veterans. Addict Behav 2016 Nov;62:47-53. [doi: 10.1016/j.addbeh.2016.06.016] [Medline: 27318948]
17. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 4;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
18. Schwarzer R, Satow L. Online intervention engagement predicts smoking cessation. Prev Med 2012 Sep;55(3):233-236. [doi: 10.1016/j.ypmed.2012.07.006] [Medline: 22813919]
19. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013 Jan 28;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
20. Balmford J, Borland R. How do smokers use a smoking cessation text messaging intervention? Nicotine Tob Res 2014 Dec;16(12):1586-1592. [doi: 10.1093/ntr/ntu111] [Medline: 25031312]
21. Saul JE, Amato MS, Cha S, Graham AL. Engagement and attrition in Internet smoking cessation interventions: insights from a cross-sectional survey of “one-hit-wonders”. Internet Interv 2016 Sep;5:23-29. [doi: 10.1016/j.invent.2016.07.001]
22. Gmel G, Kuendig H, Notari L, Gmel C. [Addiction monitoring Switzerland: Use of alcohol, tobacco and illegal drugs in Switzerland, data 2015]. Lausanne: Sucht Schweiz; 2016. URL: http://www.suchtmonitoring.ch/docs/library/gmel_thezckxvvkq4.pdf [accessed 2017-10-02] [WebCite Cache ID 6tub3toxA]
23. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
24. Ybarra ML, Prescott TL, Holtrop JS. Steps in tailoring a text messaging-based smoking cessation program for young adults. J Health Commun 2014 Dec;19(12):1393-1407. [doi: 10.1080/10810730.2014.901441] [Medline: 24766267]
25. Haug S, Castro RP, Filler A, Kowatsch T, Fleisch E, Schaub MP. Efficacy of an Internet and SMS-based integrated smoking cessation and alcohol intervention for smoking cessation in young people: study protocol of a two-arm cluster randomised controlled trial. BMC Public Health 2014 Nov 5;14:1140 [FREE Full text] [doi: 10.1186/1471-2458-14-1140] [Medline: 25369857]
26. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: a novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: overview and preliminary evaluation in the public health contexts. 2015 Presented at: Wireless Telecommunications Symposium (WTS), 2015; April 15-17, 2015; New York URL: http://cocoa.ethz.ch/downloads/2015/06/2120_WTS-2015-MobileCoach-final_2.pdf [doi: 10.1109/WTS.2015.7117255]
27. Euro.who.int. Copenhagen: WHO Regional Office for Europe; 2016. Health behaviour in school-aged children (HBSC) study: International Report from the 2013/2014 survey Internet URL: http://www.euro.who.int/__data/assets/pdf_file/0003/303438/HSBC-No.7-Growing-up-unequal-Full-Report.pdf?ua=1 [accessed 2017-10-02]
28. Collins RL, Parks GA, Marlatt GA. Social determinants of alcohol consumption: the effects of social interaction and model status on the self-administration of alcohol. J Consult Clin Psychol 1985 Apr;53(2):189-200. [Medline: 3998247]
29. Gabadinho A, Ritschard G, Müller NS, Studer M. Analyzing and visualizing state sequences in R with TraMineR. J Stat Soft 2011;40(4):1-37. [doi: 10.18637/jss.v040.i04]
30. Lê S, Josse J, Husson F. FactoMineR: an package for multivariate analysis. J Stat Soft 2008;25(1):1-18. [doi: 10.18637/jss.v025.i01]
31. Studer M. WeightedCluster library manual: a practical guide to creating typologies of trajectories in the social sciences with R. LIVES Work Pap 2013;2013(24):1-32. [doi: 10.12682/lives.2296-1658.2013.24]
32. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. New York: John Wiley & Sons; 2013.
33. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48 [FREE Full text] [doi: 10.18637/jss.v067.i01]
34. Melzer J, Rißling JK, Petermann F. [Influence of nonverbal intelligence and migration background on language abilities]. Gesundheitswesen 2015 Oct;77(10):793-798. [Medline: 26515707]
35. Ames SC, Werch CE, Ames GE, Lange LJ, Schroeder DR, Hanson AC, et al. Integrated smoking cessation and binge drinking intervention for young adults: a pilot investigation. Ann Behav Med 2010 Dec;40(3):343-349 [FREE Full text] [doi: 10.1007/s12160-010-9222-4] [Medline: 20730517]

36. Ames SC, Pokorny SB, Schroeder DR, Tan W, Werch CE. Integrated smoking cessation and binge drinking intervention for young adults: a pilot efficacy trial. Addict Behav 2014 May;39(5):848-853. [doi: 10.1016/j.addbeh.2014.02.001] [Medline: 24583274]
37. Paz Castro R, Haug S, Kowatsch T, Filler A, Schaub MP. Moderators of outcome in a technology-based intervention to prevent and reduce problem drinking among adolescents. Addict Behav 2017 Sep;72:64-71. [doi: 10.1016/j.addbeh.2017.03.013] [Medline: 28371696]
38. 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]
39. 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]
40. Whitton AE, Proudfoot J, Clarke J, Birch MR, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015 Mar 4;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]

Abbreviations

ASW: average silhouette width
CCA: complete case analysis
CPD: cigarettes smoked per day
CPW: cigarettes smoked per week
DE: decreasing engagement
HAPA: health action process approach
ITT: intention-to-treat
LOCF: last-observation-carried-forward
OM: optimal matching
OR: odds ratio
RCT: randomized controlled trial
SD: standard deviation
SE: stable engagement
SE: standard error
SNE: stable nonengagement

Edited by G Eysenbach; submitted 16.06.17; peer-reviewed by R Schwarzer, S Hermsen; comments to author 02.08.17; revised version received 07.09.17; accepted 07.09.17; published 01.11.17.

Please cite as:
Paz Castro R, Haug S, Filler A, Kowatsch T, Schaub MP
Engagement Within a Mobile Phone–Based Smoking Cessation Intervention for Adolescents and its Association With Participant Characteristics and Outcomes
J Med Internet Res 2017;19(11):e356
URL: http://www.jmir.org/2017/11/e356/
doi:10.2196/jmir.7928
PMID:29092811

Original Paper

Preventing Depression in Final Year Secondary Students: School-Based Randomized Controlled Trial

Yael Perry[1], PhD; Aliza Werner-Seidler[1], PhD; Alison Calear[2], PhD; Andrew Mackinnon[1], PhD; Catherine King[1], M Clin Psych; Jan Scott[3], MD; Sally Merry[4], PhD; Theresa Fleming[4], PhD; Karolina Stasiak[4], PhD; Helen Christensen[1], PhD; Philip J Batterham[2], PhD

[1]Black Dog Institute, University of New South Wales, Randwick, Australia

[2]National Institute for Mental Health Research, The Australian National University, Canberra, Australia

[3]Institute of Neuroscience, University of Newcastle, Newcastle upon Tyne, United Kingdom

[4]Department of Psychological Medicine, University of Auckland, Grafton, New Zealand

Corresponding Author:
Helen Christensen, PhD
Black Dog Institute
University of New South Wales
Hospital Road
Randwick,
Australia
Phone: 61 2 9382 8208
Email: h.christensen@blackdog.org.au

Abstract

Background: Depression often emerges for the first time during adolescence. There is accumulating evidence that universal depression prevention programs may have the capacity to reduce the impact of depression when delivered in the school environment.

Objective: This trial investigated the effectiveness of SPARX-R, a gamified online cognitive behavior therapy intervention for the prevention of depression relative to an attention-matched control intervention delivered to students prior to facing a significant stressor—final secondary school exams. It was hypothesized that delivering a prevention intervention in advance of a stressor would reduce depressive symptoms relative to the control group.

Methods: A cluster randomized controlled trial was conducted in 10 government schools in Sydney, Australia. Participants were 540 final year secondary students (mean 16.7 [SD 0.51] years), and clusters at the school level were randomly allocated to SPARX-R or the control intervention. Interventions were delivered weekly in 7 modules, each taking approximately 20 to 30 minutes to complete. The primary outcome was symptoms of depression as measured by the Major Depression Inventory. Intention-to-treat analyses were performed.

Results: Compared to controls, participants in the SPARX-R condition (n=242) showed significantly reduced depression symptoms relative to the control (n=298) at post-intervention (Cohen *d*=0.29) and 6 months post-baseline (*d*=0.21) but not at 18 months post-baseline (*d*=0.33).

Conclusions: This is the first trial to demonstrate a preventive effect on depressive symptoms prior to a significant and universal stressor in adolescents. It demonstrates that an online intervention delivered in advance of a stressful experience can reduce the impact of such an event on the potential development or exacerbation of depression.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12614000316606; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=365986 (Archived by WebCite at http://www.webcitation.org/6u7ou1aI9)

(J Med Internet Res 2017;19(11):e369) doi:10.2196/jmir.8241

KEYWORDS

prevention; depression; adolescent; digital cognitive behavior therapy

Introduction

Depression carries the greatest burden of disease in young people, with an annual prevalence in adolescence of 8% and approximately 80% of these cases experiencing a moderate to severe impact on functioning [1]. It has been estimated that, at most, only 36% of the burden of depression could potentially be alleviated using current knowledge and therapies [2]. Prevention represents an important alternative pathway to reducing this burden, with evidence suggesting that preventive efforts may result in a further 21% to 22% reduction [3]. A number of recent reviews and meta-analyses have found small but consistent effects for the prevention of depression in young people [4]. These reviews indicate mixed findings regarding the relative superiority of universal prevention programs (which target all individuals in a given population regardless of risk) compared to targeted programs (which are directed at high-risk individuals on the basis of the presence of existing risk factors or subthreshold symptoms). However, universal programs may offer additional advantages as they may be easier to implement, are less stigmatizing, and provide greater scope and catchment of individuals who may be on the trajectory toward developing depression but do not yet display symptoms.

The incidence of depression increases monotonically throughout adolescence [5]. However, stressful life events are precursors of and possible causal factors for the onset of depression [6]. Final school examinations are a significant stressor for most adolescents, with exams and academic outcomes endorsed most strongly among school-related stressors [7]. In the Australian context, these final exams determine tertiary education entrance scores, rendering their outcome critically important. More than 40% of final year secondary students report elevated symptoms of depression, anxiety, and stress [8]. In extreme cases, exam stress has been linked to suicidal ideation, behavior, and completion [9]. To date, the potential for prevention has not been examined in the context of school-based stressors. Schools are an ideal location in which to deliver mental health interventions that address these difficulties, not only because young people spend more time in these institutions than any other, but also because of the integral role schools play in students' social, academic, cognitive, emotional, and behavioral development. While school-based prevention programs have typically taken the format of face-to-face, group-based interventions, online interventions in the school setting are becoming increasingly appealing [10]. Automated online programs guarantee fidelity, since the information cannot be distorted, diluted, or contaminated during delivery [11]. Consequently, implementation of an online universal depression prevention program in the context of a stressor widely experienced in the target sample represented a unique opportunity to evaluate the effectiveness of such an approach.

Our cluster randomized controlled trial (RCT) was designed to test the effectiveness of SPARX-R, an online, cognitive behavioral gamified intervention, in preventing the development or exacerbation of depressive symptoms in final year secondary students relative to an online active control program (lifeSTYLE). SPARX was originally designed as a treatment intervention for depression; however, it was subsequently adapted for use as a universal preventive intervention, known as SPARX-R. SPARX has been shown to result in a decline in symptoms of depression and improved remission rates in a help-seeking sample of adolescents and was found to be non-inferior to treatment as usual (consisting primarily of face-to-face therapy with a counselor, general practitioner, or clinical psychologist) [12]. However, in a recent school trial for individuals with subthreshold depression, SPARX failed to demonstrate effects superior to a control condition despite having a positive impact on depressive symptoms [13]. The authors suggest that the absence of an intervention effect may have been due to the effectiveness of the control condition, which essentially operated as a self-monitoring intervention and may have been sufficient to improve those with mild depressive symptoms. The effectiveness of self-monitoring has been demonstrated in other trials (see, for example, Kauer et al [14]). In our trial, it was predicted that students assigned to the SPARX-R condition would report lower levels of depressive symptoms immediately following the intervention and at 6-month follow-up relative to participants in the attention control condition.

As well as evaluating the impact of this novel intervention on depressive symptoms, this study also examined the effect of SPARX-R on secondary outcomes including anxiety, stigma, academic performance, and suicidal ideation. Given the substantial comorbidity of anxiety and depression and the fact that anxiety symptoms can decline in the context of some depression prevention programs [15], it was expected that participants in the SPARX-R group would also experience reduced anxiety. There is evidence that increased knowledge regarding depression is associated with reduced stigma [16]. Therefore, it was predicted that stigma would also be decreased in the SPARX-R group. Outcomes of final examinations for participants were sought from schools to evaluate whether any impact of the intervention flowed through to academic performance, an expectation arising from the association of depression with poor academic outcomes [17]. Although depression and suicidality are associated [18], they may not follow the same trajectory and so the examination of outcomes pertaining to suicidal ideation and behavior was exploratory.

Methods

Study Design

A cluster RCT was conducted with 2 parallel arms consisting of an experimental condition (SPARX-R) and an attention-matched control condition (lifeSTYLE). The trial was designed for implementation in schools in accordance with the Australian academic calendar: baseline assessments were conducted at the start of the academic year (February), followed by a 5-week intervention phase and post-intervention assessments. Six-month follow-up assessments were conducted midyear, just prior to a significant period of academic assessments that contribute to students' final secondary school grade (the exam period). The final follow-up assessment was conducted 18 months after baseline. At this time, participants had graduated from high school, and the majority were engaged in post-secondary studies. This trial received ethics approval

from the New South Wales Department of Education and Training (SERAP2014001), the University of New South Wales (HC14105), and the Australian National University (2014/325). The study protocol is published [19].

Participants

The trial was conducted in selective and partially selective government secondary schools in metropolitan Sydney. Selective schools use an entry exam to enroll students with superior academic ability, while partially selective schools offer both selective and comprehensive (non-selective) streams. Academically gifted students were selected because final school exam outcomes are critical to them. They are particularly sensitive to the stressful nature of exams, with studies finding they experience adverse stress reactions to academic failures coupled with pressure to succeed from parents, schools, the media, and students themselves [20]. The academic workload and considerable number of hours spent studying in a demanding educational program is likely to compound this pressure and act as an additional stressor [21]. These attributes made this group a particularly relevant sample in which to test the stress-diathesis model of depression.

All adolescents enrolled in their final year of secondary school in participating schools were invited to participate in the trial. All students at participating partially selective schools were eligible to participate in the trial; however, streaming status (ie, selective or comprehensive) was recorded and evaluated in analyses. Due to the universal nature of the study, there were no exclusion criteria. Written, informed consent was sought from students and their parents prior to the start of the trial. Students who did not provide consent were able to access the assigned intervention; however, only those with consent completed the research questionnaires.

Interventions

SPARX-R

SPARX-R is a revised version of SPARX, which was developed as an unguided, interactive program using the format of a fantasy game providing cognitive behavioral skills to treat mild to moderate symptoms of depression in help-seeking adolescents.

SPARX-R provides users with the same skills as those in SPARX [12]; however, the revised version is framed in preventive terms. For example, participants are told that "this version of SPARX was made to help young people who are having hassles and feeling down, stressed, or angry a lot of the time. Even if you are doing fine, SPARX-R can help strengthen your skills for dealing with problems when they do come along." Further, for the version of SPARX-R used, care was taken to ensure that terminology and local helplines and services were suitable for Australia. SPARX-R users choose and personalize an avatar and are led through the program by a virtual guide who provides context and links the content of the program to their real-life experiences. The user navigates their way through a series of challenges within a fantasy world that has been overrun by GNATs (gloomy, negative, automatic thoughts), with the mission of restoring balance in the game world. The program has 7 modules (levels), each of which takes approximately 20 to 30 minutes to complete. Participating schools scheduled curriculum time for participants to complete 1 to 2 modules per week, allowing a few days in between modules for students to process what they learned and practice new skills before beginning the next module. The intervention was completed over the course of 5 to 7 weeks in class under teacher supervision. The modules cover the following topics: finding hope, being active, dealing with strong emotions, overcoming problems, recognizing unhelpful thoughts, challenging unhelpful thoughts, and bringing it all together. Key skills taught by the program were relaxation, activity scheduling and behavioral activation, emotion regulation, interpersonal skills, problem solving, cognitive restructuring, and distress tolerance [12]. SPARX-R was delivered to students on desktop computers via the Internet in school classrooms and supplemented with a paper notebook for students to review key messages from each module and record personal comments. A trailer for how SPARX works can be found at sparx.org.nz.

lifeSTYLE

lifeSTYLE is an adaptation of an interactive, online program originally developed as a control intervention for a trial targeting adults with suicidal thoughts [22]. The format and structure of the program was retained, but the content was adapted to suit adolescents. This active comparison was used to control for the large placebo effect commonly seen in depression prevention and treatment trials, which has been cited as a key methodological weakness in many studies. The aim of the intervention was to provide an engaging and useful resource for young people that matched the intervention in terms of duration and attention without providing any direct mental health content. As with SPARX, lifeSTYLE consisted of 7 modules, each of which took 20 to 30 minutes to complete. The program covers the following topics: independence, participating in your community, work skills, mobile phone safety and hygiene, healthy skin, sustainable eating, and maintaining a healthy home environment. Each module includes information about the specified topic as well as interactive activities such as quizzes, mythbusters, videos, and scenarios to which students can respond. As with SPARX-R, the intervention was delivered online to students in school classrooms.

Randomization and Masking

Each participating school (cluster) was randomized to SPARX-R or lifeSTYLE by a statistician not involved in the implementation of the trial and blind to the identity of the schools. All study staff remained blind to condition except for the 2 trial managers involved in the communication of the materials and instructions to the schools (YP, AWS). The method of Carter and Hood [23] sought to ensure balance between arms. Balance variables were gender, number of enrolled students, Index of Community Socio-Educational Advantage for each school, and language background other than English.

No research personnel were directly involved in the delivery of the prevention interventions. Schools were not informed whether their assigned program represented the experimental or control condition. All outcome measures were self-report and were completed privately via an online portal. The trial statistician was blind to school allocation.

Procedure

Following school-level randomization, participants completed baseline assessments and their allocated intervention modules. Program sessions were scheduled at each school's convenience. For the most part, students collectively completed the sessions during the first 20 to 30 minutes of a designated class period and then resumed their regular lessons for the remainder of the period. The majority of participants completed their post-intervention assessments in scheduled class sessions within 1 week of the intervention period. A follow-up assessment session was scheduled approximately 6 months after the baseline assessment (July; approximately 2 weeks prior to the exam period). A final follow-up assessment was scheduled 18 months after the baseline assessment; approximately 7 to 8 months after participants had completed their final exams. Exam results were provided by school administrators approximately 6 weeks after the final exam period ended (November-December).

Outcome Measures

Primary Outcome Measure

The Major Depression Inventory (MDI) is a 12-item self-report measure of depressive symptoms [24]. The items of the MDI evaluate the presence and duration of depressive symptoms according to both *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*, (ICD-10) and *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*, (DSM-IV) criteria. Respondents rate the degree to which they have experienced each of 10 symptoms over the preceding 2 weeks on a 6-point scale, ranging from 0 (at no time) to 5 (all of the time). The MDI has the following cut-off points: 21 to 25 for mild depression, 26 to 30 for moderate depression, and 31 to 50 for severe depression [25]. The MDI has acceptable sensitivity and specificity for the diagnosis of depression according to the ICD-10 and DSM-IV [24]. In this trial, Cronbach alpha, assessed at baseline, was .88.

Secondary Outcome Measures

The Spence Children's Anxiety Scale (SCAS) is a 44-item measure comprising 6 subscales. The scale was designed to measure the severity of children and adolescents' anxiety symptoms based broadly on DSM-IV criteria for anxiety disorders [26]. Respondents rate the degree to which they experience each symptom on a 4-point frequency scale, ranging from 0 (never) to 3 (always). Only items on the social phobia and generalized anxiety subscales were administered. The scale has demonstrated high reliability [26]. In this trial, Cronbach alphas for the social phobia and generalized anxiety scales were .76 and .81, respectively.

Three items from the Youth Risk Behavior Survey concerning suicidal thoughts, plans, and attempts over the previous month were used to assess suicidality. Endorsement of any of these items triggered the trial risk management protocol. Studies have shown that the suicidality items demonstrate reliability [27].

The Depression Stigma Scale (DSS) is an 18-item measure that assesses personal and perceived stigma toward depression. In this study, only the 9-item personal stigma subscale was used, as this subscale was most relevant. Items require the participant to rate how strongly they agree with a statement about depression (eg, "people with depression are unpredictable") on a 5-point Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree). The sum of each of the items yields a total stigma score, where higher scores indicate greater stigma. For this sample, Cronbach alpha was .77.

Final exam results were also collected and standardized using the Australian Tertiary Admission Rank such that the 2 groups could be compared on academic outcomes.

Sample Size

Based on previous depression prevention research using an online intervention [15], a post-intervention Cohen *d* effect size for the primary outcome measure of .20 was used for sample size calculations. Power was set at 80%, alpha=.05 (2-tailed), and a correlation of .5 assumed between baseline and endpoint scores. To allow for possible clustering effects, a design effect was calculated assuming an intraclass correlation (ICC) of .02 and an average class size of 25. The estimate of the ICC was derived from a previous Australian school-based study that found a nonsignificant ICC of .02 [15]. The estimated sample size was 1166. The total target size sample size was set at 1600, 800 students per condition, in order to accommodate possible attrition at a rate up to 20% [28].

The protocol [19] specified 2 stages of recruitment over consecutive years, with a planned analysis conducted using data from the first cohort as soon as available. In addition to evaluating effectiveness of the intervention, a review of first cohort outcomes was planned, aimed at identifying any refinement of content and mode of delivery required before proceeding to the second stage of recruitment. Schools recruited to the first cohort had an enrollment of 1677 eligible students. Under the assumptions outlined above, for this cohort alone, there was 80% power to detect an effect size of 0.40 standard deviations between groups.

Data Analysis

Primary analyses were undertaken on an intent-to-treat (ITT) basis. Effectiveness of SPARX-R was established using a planned contrast of change from baseline to post-intervention in the active compared to placebo condition on the MDI using a mixed-model repeated measures (MMRM) analysis that incorporates all available data under the missing-at-random assumption [29]. School was included in analyses as a random effect to evaluate and accommodate clustering effects. Variables used in determining allocation balance were evaluated and retained in 2 sensitivity analyses, 1 adjusting for characteristics differing at baseline and 1 excluding non-selective students. An unconstrained variance-covariance matrix was used to model within-individual dependencies, and degrees of freedom were estimated using Satterthwaite's method [30]. Between-group effect sizes (Cohen *d*) were calculated using observed mean loss/gain scores from pre-intervention to post-intervention, from pre-intervention to 6-month follow-up, and from pre-intervention to 18-month follow-up, divided by pooled standard deviation at the later time-point. Positive effect sizes indicate changes in favor of the intervention condition.

Although our MMRM analytic approach was conducted under the robust assumption that data were missing-at-random, we also conducted a completer analysis, excluding participants who did not complete a follow-up assessment. This analysis was conducted to provide evidence for whether findings were consistent across alternative assumptions about missingness. Due to high attrition from the trial at follow-up, consistency across multiple analytical approaches may provide added confidence in the findings. Outcomes of the completer analysis are briefly summarized with comparison to the ITT analyses, as completer analyses may be more prone to bias [31].

Prevalence rates of major depressive disorder and suicidal ideation between the 2 treatment arms at post-intervention, 6 months, and 18 months were compared using Fisher exact test. Analyses of secondary variables also used MMRM methods. Subsidiary analyses of students who completed post-test evaluations (protocol compliers) were used to estimate the efficacy of SPARX-R in participants who completed sufficient modules from the program to have a potential clinical impact (estimated to be 4 or more modules) [12]. Separate analyses for completer subgroups (eg, completed 0 to 3 modules vs 4 to 7 modules) were also undertaken, and clinically significant change (improvement and decline) was assessed in program completers. Due to scheduling difficulties, 2 participants in lifeSTYLE and 22 in SPARX-R completed a 6-month assessment after completion of exams. As the stressor of interest had consequently passed, the 6-month data from these participants were excluded from analyses. All analyses were undertaken using SPSS version 23 (IBM Corp). The trial was registered with the Australian New Zealand Clinical Trials Registry [ANZCTRN12614000316606].

Results

A total of 23 secondary schools were approached between August 1 and November 28, 2014, to participate in the trial. Of 14 schools who initially agreed to participate, 4 dropped out prior to randomization leaving 10 schools contributing 7 to 126 participants each. A total of 540 students provided personal consent and obtained parental consent to participate. The Consolidated Standards of Reporting Trials (CONSORT) flowchart outlines recruitment, randomization, and participation (see Figure 1). Baseline assessments commenced in February 2015 with 6-month follow-up assessments completed by August 2015 and 18-month follow-up assessments completed by August 2016.

Sample characteristics are shown in Table 1. There were no differences between the SPARX-R and lifeSTYLE conditions on the basis of country of birth, language primarily spoken at home, age, depression symptoms, generalized anxiety disorder symptoms, social phobia symptoms, or subjective health. However, males, students from selective programs, and students living with both parents together had significantly greater representation in the SPARX-R condition relative to lifeSTYLE. Participants in SPARX-R completed fewer intervention modules than those in lifeSTYLE. Estimated marginal means and standard errors were derived from models fitted for primary and secondary continuous outcomes at each time point by condition (see Table 2).

Table 1. Sample characteristics for the SPARX-R and lifeSTYLE conditions.

Characteristic		SPARX-R n=242 n (%)	lifeSTYLE n=298 n (%)	Chi-square	*P* value
School program					
	Non-selective	4 (1.7)	72 (24.2)	55.9	<.001
	Selective	238 (98.3)	226 (75.8)		
Sex					
	Male	117 (48.3)	82 (27.5)	24.9	<.001
	Female	125 (51.7)	216 (72.5)		
Country of birth					
	Australia	195 (80.6)	236 (79.2)	0.2	.69
	Other	47 (19.4)	62 (20.8)		
Language spoken at home					
	English	132 (54.5)	175 (58.7)	1.0	.33
	Other	110 (45.4)	123 (41.3)		
Live with both parents together					
	Yes	216 (89.3)	235 (78.9)	10.5	<.001
	No	26 (10.7)	63 (21.1)		

Table 2. Mean and standard error for continuous outcomes at each time point (estimated from mixed effect models).

Test	SPARX-R				lifeSTYLE			
	Baseline n=242 mean (SE)	Post n=206 mean (SE)	6 months n=140 mean (SE)	18 months n=40 mean (SE)	Baseline n=298 mean (SE)	Post n=200 mean (SE)	6 months n=201 mean (SE)	18 months n=64 mean (SE)
MDI[a]	14.9 (0.9)	11.9 (0.9)	13.3 (1.0)	10.0 (1.1)	14.4 (0.9)	14.7 (0.9)	15.3 (0.9)	11.9 (1.0)
SCAS GAD[b]	6.7 (0.4)	5.9 (0.4)	6.5 (0.6)	5.1 (0.5)	6.8 (0.4)	6.4 (0.4)	6.6 (0.4)	5.7 (0.5)
SCAS SA[c]	7.2 (0.5)	6.4 (0.5)	6.5 (0.5)	6.1 (0.6)	7.4 (0.4)	6.8 (0.5)	6.6 (0.5)	6.5 (0.5)
DSS[d]	9.4 (1.0)	8.9 (1.1)	8.4 (1.1)	8.1 (1.1)	8.8 (1.0)	8.8 (1.0)	8.1 (1.0)	7.8 (1.0)

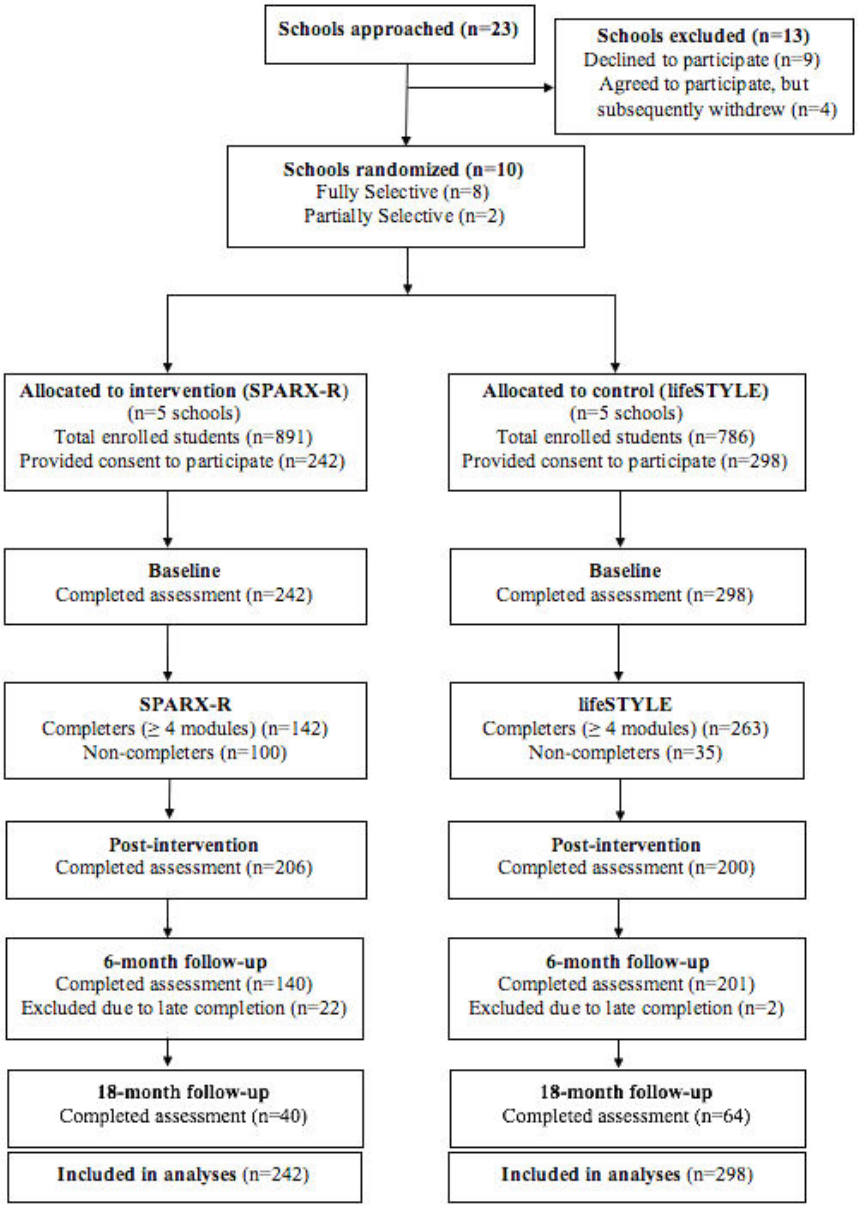
[a]MDI: Major Depression Inventory.

[b]SCAS GAD: Spence Child Anxiety Scale–Generalized Anxiety Disorder.

[c]SCAS SA: Spence Child Anxiety Scale–Social Anxiety.

[d]DSS: Depression Stigma Scale.

Figure 1. CONSORT Diagram.



The ICC for depressive symptoms was 0.017. Primary analyses showed that participants in the SPARX-R condition had a significantly greater reduction in MDI scores than those in the lifeSTYLE condition both post-intervention ($t_{(416.1)}$=−4.78, P<.001) and at 6-month follow-up ($t_{(317.6)}$=−2.76, P=.01). The effect was not significant at 18-month follow-up ($t_{(132.1)}$=−1.82, P=.07). Effect sizes were small post-intervention (Cohen d=0.29, 95% CI 0.09 to 0.49), at 6-month follow-up (Cohen d=0.21, 95% CI −0.01 to 0.42), and 18-month follow-up (Cohen d=0.33, 95% CI −0.06 to 0.73).

When completer analyses were conducted, the effects of the intervention on depression scores remained significant ($F_{(3, 254.0)}$=8.00, P<.001), with significant effects at post-test ($t_{(410.4)}$=−4.84, P<.001) and 6-month follow-up ($t_{(368.1)}$=−2.82, P=.005) but not at 18 months ($t_{(132.4)}$=−1.93, P=.055), consistent with the ITT analysis. Secondary outcomes also remained nonsignificant in completer analyses.

For participants who completed 4 or more modules, the intervention was effective at both post-intervention ($t_{(328.3)}$=−4.55, P<.001) and 6 months ($t_{(313.6)}$=−2.68, P=.01) but not at 18-month follow-up ($t_{(116.4)}$= −1.54, P>.05). Conversely, those who completed fewer modules showed no significant intervention effect (P=.83 post-intervention, P=.57 at 6 months) except for at 18 months ($t_{(15.5)}$= −3.08, P=.01). A clinically significant improvement in depression, as indicated by the Reliable Change Index [32], was found in a larger proportion of completers of SPARX-R than controls at post-test (65/206, 31.6%, vs 26/200, 13.0%, P<.001) and 6-month follow-up (35/126, 27.8%, vs 33/202, 16.3%, P=.02) but not at 18-month follow-up (15/41, 37%, vs 17/64, 27%, P=.29).

No significant effects were found on any of the secondary outcomes. Generalized anxiety and social anxiety were significantly reduced at post-intervention relative to baseline for both groups, but there was no significant condition by time interaction effect for either anxiety outcome. There were no differences between conditions in the prevalence of depression at any time, based on MDI caseness criteria [24], with caseness of 4.9% (10/206) in SPARX-R and 8.5% (17/200) in lifeSTYLE at post-intervention (P=.17), 11.9% (15/126) and 9.4% (19/202), respectively, at 6 months (P=.46), and 2.4% (1/41) and 4.7% (3/64), respectively, at 18 months (P>.99). Few participants reported suicidal ideation at any time point, and no significant differences were found between groups: 3.4% (7/205) in SPARX-R and 4.1% (8/196) in lifeSTYLE at post-intervention (P=.80), 2.4% (3/126) and 4.0% (8/198), respectively, at 6 months (P=.54), and 3% (1/39) and 5% (3/62), respectively, at 18 months (P>.99). Notably, academic outcomes did not differ between the 2 groups (P=.41).

Discussion

Principal Findings

The online gamified prevention intervention, SPARX-R, was effective at reducing depression symptoms in secondary students prior to their final exams. The effect size was small but robust and corresponds to previous results from cognitive behavior therapy (CBT)–based school programs delivered face-to-face based on recent meta-analyses [4]. Effects remained significant after controlling for selective streaming status as well as characteristics found to differ substantially at baseline. Completion of the program was associated with clearly improved primary outcomes, such that SPARX-R completers reported fewer depressive symptoms postintervention and at 6-month follow-up than their lifeSTYLE counterparts, while no intervention effect was detected for noncompleters. Differences in rates of clinically significant change were demonstrated, implying that the intervention produces meaningful change in depressive symptoms under universal administration. Given that a small effect at the population level can have a substantial impact on overall mental health, we assess the impact of the magnitude of effect of the intervention as promising. This is an important finding as it highlights the effectiveness of SPARX-R in the context of a truly universal sample comprising those who are at elevated risk for developing depression and those who are not.

No intervention effects were found for generalized anxiety, a finding at variance with the use of SPARX as a treatment intervention in the initial study [12], although anxiety was not measured in the second SPARX study [13]. The overall reduction in anxiety observed in the course of the trial is somewhat surprising as it might be expected to increase with the approach of final exams.

No between-group differences were found in depression caseness, suicidality, or academic outcomes. Low base rates for caseness and suicidality would likely have obscured any effects that existed. It is unclear why decreased symptoms did not improve academic performance, (although findings were in the expected direction), but it is possible that due to the high achieving nature of the sample there was insufficient variability in academic performance to detect change or students may have already been performing at ceiling. This was also a universal group of relatively psychologically well students, so differences may be more likely to emerge in a clinical sample or at least at the more severe end of the spectrum [17]. This question will need to be addressed in future prevention research.

Strengths and Limitations

A key strength of the study is its implementation in school settings where the intervention will ultimately be delivered, directly supporting the translational capacity of this research into practice. The trial is noteworthy having been mounted in a sample faced with a meaningful, real-world stressor. Indeed, we believe that the anticipation of the stressor may have led to greater engagement, addressing a problem that has been identified in more recent digital trials [33]. Thus, the significant effects detected are of great potential importance, given that they were obtained under pragmatic, real-world conditions. The use of an active attention control condition, which was matched in terms of length, delivery format, and interactivity, means the effects are attributable to the intervention rather than involvement in a research trial or generic placebo effects.

Several limitations associated with this study warrant mention. In common with most Internet-delivered universal interventions, completion rates of the active intervention were modest (59%

in the SPARX-R group; defined as completion of 4 or more of 7 modules) although substantially higher than comparable studies in secondary schools [15]. While this rate is lower than the 86% completion rate observed in the SPARX treatment trial [12], it is important to note that this sample was universal rather than help-seeking, which may account for differences in motivation. Completion may have also been compounded by technical difficulties. Several technical problems occurred during implementation (particularly in schools allocated to the SPARX-R intervention), which impacted some students' ability to complete all of their allocated modules during class time. This was largely due to the excessive load on the school information technology (IT) system associated with multiple students simultaneously accessing the online research platform and downloading the SPARX-R game files. Indeed, the completion rate of 4 or more modules in the lifeSTYLE group (which did not encounter the same degree of difficulty due to a smaller load on IT systems) was substantially higher (88%). In response to these issues, the stability and download size of the SPARX-R program has been reviewed with the assistance of software developers. A mobile app of the intervention is also being developed and will be tested in the future.

Our decision to recruit in academically selective schools means that caution must be exercised in generalizing results to the broader school population. However, approximately 80% of our sample was born outside Australia and over half did not speak English at home, suggesting this was a culturally, if not academically, diverse sample.

Resources and feasible sample size precluded mounting a trial with incidence of clinically diagnosed depressive disorder as its primary outcome. This limitation is shared by nearly all such school-based trials and might be precluded in any case due to the demands of the education curriculum in the last year of high school. Accordingly, this trial aimed to maximize ecological validity in order to determine the true effect of a prevention program under real-world conditions, although we acknowledge that a small proportion of participants may have met criteria for the disorder at baseline.

Prior Research

To date, the majority of school-based depression prevention trials have delivered material face-to-face in groups. Five published studies have examined the effectiveness of online prevention interventions in the school setting. Calear and colleagues [15] were the first to demonstrate the utility of an online CBT program (MoodGYM) in high school students, demonstrating a positive effect for the program at post-intervention and 6-month follow-up for anxiety (Cohen d=0.15 and 0.25, respectively) and depressive symptoms (in males only; d=0.43 and 0.31, respectively). Wong and colleagues [34] conducted an RCT investigating 2 separate, blended online/offline CBT interventions (Thiswayup Schools) targeting depression and anxiety compared to regular health classes. Students who received the depression intervention reported lower symptoms of both depression (d=0.14) and anxiety (d=0.29) at post-intervention. More recently, 2 trials have not shown significant effects, findings that have been interpreted as a failure to engage or attributed to the strength of the control group intervention [13,33]. Our study suggests that the context in which the intervention is offered must also be taken into consideration. Adult findings indicate the power of preparatory CBT in the context of a stressor. For example, in 1 study doctors in training were randomly assigned to 2 study groups, either MoodGYM or an attention control group, prior to taking up their medical internship, a period of high stress [35]. Results showed that suicide ideation was less frequently endorsed in the intervention group than the control over the course of the year. Therefore, interventions delivered prior to major transitions or major stressors may provide greater engagement (both from the students and the school administrators) and be more effective than those delivered at arbitrary times.

SPARX-R can be delivered without direct facilitation and requires only minimal supervision from teachers, reducing the human resources required to deliver face-to-face school-based programs, and may offer an effective and scalable universal prevention approach for students. Indeed, it could be the first step in a stepped care approach to mental health interventions, whereby young people within the school setting are provided with access to a universal, skills-oriented intervention, with those who require additional support being stepped up to more intensive care as needed.

Conclusion

This is the first trial to demonstrate a preventive effect on depression prior to a significant and universal stressor in adolescents. It extends the evidence base for appropriately developed CBT skills delivered through a gamified program as being broadly effective for young people regardless of whether their symptoms are minimal, elevated, or meet diagnostic criteria. Our results highlight the potential utility of delivering prevention interventions in advance of a key stressor and the program feasibility and effectiveness of using an engaging interactive tool for depression prevention. Major implementation studies of these types of digital interventions are now warranted.

Acknowledgments

This project was supported by an Australian National Health and Medical Research Council (NHMRC) project grant 1061072 to HC. HC is supported by an NHMRC John Cade fellowship (1056964). ALC and PJB are supported by NHMRC fellowships 1013199 and 1083311. The funders had no role in the design, execution, analyses, or data interpretation for this study or in the decision to submit the results for publication. We would like to thank Professor Julio Lucinio and Dr Tara Donker for assistance in developing the research concept and Noel Thomsen for assistance and guidance in school recruitment. We would also like to thank the schools that participated in our trial and, in particular, the teachers and school counselors who facilitated its

implementation. We would also like to acknowledge the work of the systems administrators at the Black Dog Institute, who designed and implemented the online platform that delivered both the interventions and assessments to trial participants.

Conflicts of Interest

YP, AWS, ALC, AM, CK, JS, HC, and PB declare no conflicts of interest. The intellectual property rights for SPARX are owned by UniServices Ltd at the University of Auckland. In the event of commercialization, SM, TF, and KS, as developers of SPARX, could derive financial benefit from its operation.

References

1. Lawrence D, Johnson S, Hafekost J, Boterhoven DHK, Sawyer M, Ainley J. The mental health of children and adolescents: report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing. 2015. URL: http://www.health.gov.au/internet/main/publishing.nsf/content/9DA8CA21306FE6EDCA257E2700016945/$File/child2.pdf [accessed 2017-10-12] [WebCite Cache ID 6uAFMa4hC]
2. Andrews G, Issakidis C, Sanderson K, Corry J, Lapsley H. Utilising survey data to inform public policy: comparison of the cost-effectiveness of treatment of ten mental disorders. Br J Psychiatry 2004 Jun;184:526-533 [FREE Full text] [Medline: 15172947]
3. van Zoonen K, Buntrock C, Ebert DD, Smit F, Reynolds CF, Beekman ATF, et al. Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. Int J Epidemiol 2014 Apr;43(2):318-329 [FREE Full text] [doi: 10.1093/ije/dyt175] [Medline: 24760873]
4. Werner-Seidler A, Perry Y, Calear AL, Newby JM, Christensen H. School-based depression and anxiety prevention programs for young people: a systematic review and meta-analysis. Clin Psychol Rev 2017 Feb;51:30-47 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.005] [Medline: 27821267]
5. Merikangas KR, Cui L, Kattan G, Carlson GA, Youngstrom EA, Angst J. Mania with and without depression in a community sample of US adolescents. Arch Gen Psychiatry 2012 Sep;69(9):943-951. [doi: 10.1001/archgenpsychiatry.2012.38] [Medline: 22566563]
6. Kendler KS, Kuhn JW, Vittum J, Prescott CA, Riley B. The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: a replication. Arch Gen Psychiatry 2005 May;62(5):529-535. [doi: 10.1001/archpsyc.62.5.529] [Medline: 15867106]
7. Kouzma NM, Kennedy GA. Self-reported sources of stress in senior high school students. Psychol Rep 2004 Feb;94(1):314-316. [doi: 10.2466/pr0.94.1.314-316] [Medline: 15077784]
8. Smith L, Sinclair K. Transforming the HSC: affective implications. Change Transform Educ 2000;3(2):67-79.
9. Rubenstein JL, Halton A, Kasten L, Rubin C, Stechler G. Suicidal behavior in adolescents: stress and protection in different family contexts. Am J Orthopsychiatry 1998 Apr;68(2):274-284. [Medline: 9589765]
10. Hetrick SE, Cox GR, Witt KG, Bir JJ, Merry SN. Cognitive behavioural therapy (CBT), third-wave CBT and interpersonal therapy (IPT) based interventions for preventing depression in children and adolescents. Cochrane Database Syst Rev 2016 Aug 09(8):CD003380. [doi: 10.1002/14651858.CD003380.pub4] [Medline: 27501438]
11. Christensen H, Griffiths KM. The prevention of depression using the Internet. Med J Aust 2002 Oct 7;177 Suppl:S122-S125. [Medline: 12358571]
12. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012;344:e2598 [FREE Full text] [Medline: 22517917]
13. Poppelaars M, Tak YR, Lichtwarck-Aschoff A, Engels RCME, Lobel A, Merry SN, et al. A randomized controlled trial comparing two cognitive-behavioral programs for adolescent girls with subclinical depression: a school-based program (Op Volle Kracht) and a computerized program (SPARX). Behav Res Ther 2016 May;80:33-42. [doi: 10.1016/j.brat.2016.03.005] [Medline: 27019280]
14. Kauer SD, Reid SC, Crooke AHD, Khor A, Hearps SJC, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
15. Calear AL, Christensen H, Mackinnon A, Griffiths KM, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]
16. Perry Y, Petrie K, Buckley H, Cavanagh L, Clarke D, Winslade M, et al. Effects of a classroom-based educational resource on adolescent mental health literacy: a cluster randomized controlled trial. J Adolesc 2014 Oct;37(7):1143-1151. [doi: 10.1016/j.adolescence.2014.08.001] [Medline: 25151646]
17. Fröjd SA, Nissinen ES, Pelkonen MUI, Marttunen MJ, Koivisto A, Kaltiala-Heino R. Depression and school performance in middle adolescent boys and girls. J Adolesc 2008 Aug;31(4):485-498. [doi: 10.1016/j.adolescence.2007.08.006] [Medline: 17949806]

18. Bruffaerts R, Kessler RC, Demyttenaere K, Bonnewyn A, Nock MK. Examination of the population attributable risk of different risk factor domains for suicidal thoughts and behaviors. J Affect Disord 2015 Nov 15;187:66-72. [doi: 10.1016/j.jad.2015.07.042] [Medline: 26321257]
19. Perry Y, Calear AL, Mackinnon A, Batterham PJ, Licinio J, King C, et al. Trial for the Prevention of Depression (TriPoD) in final-year secondary students: study protocol for a cluster randomised controlled trial. Trials 2015 Oct 12;16:451 [FREE Full text] [doi: 10.1186/s13063-015-0929-1] [Medline: 26458896]
20. Suldo SM, Shaunessy E, Hardesty R. Relationships among stress, coping, and mental health in high-achieving high school students. Psychol Schs 2008 Apr;45(4):273-290. [doi: 10.1002/pits.20300]
21. Kouzma NM, Kennedy GA. Homework, stress, and mood disturbance in senior high school students. Psychol Rep 2002 Aug;91(1):193-198. [doi: 10.2466/pr0.2002.91.1.193] [Medline: 12353781]
22. van Spijker BAJ, Calear AL, Batterham PJ, Mackinnon AJ, Gosling JA, Kerkhof AJFM, et al. Reducing suicidal thoughts in the Australian general population through web-based self-help: study protocol for a randomized controlled trial. Trials 2015;16:62 [FREE Full text] [doi: 10.1186/s13063-015-0589-1] [Medline: 25886941]
23. Carter BR, Hood K. Balance algorithm for cluster randomized trials. BMC Med Res Methodol 2008 Oct 09;8:65 [FREE Full text] [doi: 10.1186/1471-2288-8-65] [Medline: 18844993]
24. Bech P, Rasmussen NA, Olsen LR, Noerholm V, Abildgaard W. The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. J Affect Disord 2001 Oct;66(2-3):159-164. [Medline: 11578668]
25. Olsen LR, Jensen DV, Noerholm V, Martiny K, Bech P. The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. Psychol Med 2003 Feb;33(2):351-356. [Medline: 12622314]
26. Spence SH. A measure of anxiety symptoms among children. Behav Res Ther 1998 May;36(5):545-566. [Medline: 9648330]
27. Brener ND, Kann L, McManus T, Kinchen SA, Sundberg EC, Ross JG. Reliability of the 1999 youth risk behavior survey questionnaire. J Adolesc Health 2002 Oct;31(4):336-342. [Medline: 12359379]
28. Christensen H, Griffiths KM, Farrer L. Adherence in Internet interventions for anxiety and depression: systematic review. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
29. Verbeke G, Molenberghs G. Pattern-mixture models. In: Linear Mixed Models for Longitudinal Data. New York: Springer; 2010:275-293.
30. Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics Bull 1946 Dec;2(6):110-114. [doi: 10.2307/3002019]
31. Salim A, Mackinnon A, Christensen H, Griffiths K. Comparison of data analysis strategies for intent-to-treat analysis in pre-test-post-test designs with substantial dropout rates. Psych Res 2008;160(3):335-345. [doi: 10.1016/j.psychres.2007.08.005]
32. 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]
33. Calear AL, Batterham PJ, Poyser CT, Mackinnon AJ, Griffiths KM, Christensen H. Cluster randomised controlled trial of the e-couch Anxiety and Worry program in schools. J Affect Disord 2016 May 15;196:210-217. [doi: 10.1016/j.jad.2016.02.049] [Medline: 26926660]
34. Wong N, Kady L, Mewton L, Sunderland M, Andrews G. Preventing anxiety and depression in adolescents: a randomised controlled trial of two school based Internet-delivered cognitive behavioural therapy programmes. Internet Interv 2014 Apr;1(2):90-94. [doi: 10.1016/j.invent.2014.05.004]
35. Guille C, Zhao Z, Krystal J, Nichols B, Brady K, Sen S. Web-based cognitive behavioral therapy intervention for the prevention of suicidal ideation in medical interns: a randomized clinical trial. JAMA Psychiatry 2015 Dec;72(12):1192-1198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1880] [Medline: 26535958]

Abbreviations

ANZCTR: Australian and New Zealand Clinical Trials Registry
CBT: cognitive behavior therapy
CONSORT: Consolidated Standards of Reporting Trials
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
DSS: Depression Stigma Scale
GNAT: gloomy, negative, automatic thought
ICC: intraclass correlation
ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10thRevision
IT: information technology
ITT: intention-to-treat
MDI: Major Depression Inventory
MMRM: mixed-model repeated measures
NHMRC: National Health and Medical Research Council

SCAS: Spence Child Anxiety Scale
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 19.06.17; peer-reviewed by C Buntrock, Q Yuan; comments to author 27.07.17; revised version received 08.08.17; accepted 20.09.17; published 02.11.17.

Please cite as:
Perry Y, Werner-Seidler A, Calear A, Mackinnon A, King C, Scott J, Merry S, Fleming T, Stasiak K, Christensen H, Batterham PJ
Preventing Depression in Final Year Secondary Students: School-Based Randomized Controlled Trial
J Med Internet Res 2017;19(11):e369
URL: http://www.jmir.org/2017/11/e369/
doi:10.2196/jmir.8241
PMID:29097357

Original Paper

Sexual Health and the Internet: Cross-Sectional Study of Online Preferences Among Adolescents

Antonella Juline von Rosen[1,2], MA, MBBS; Frederik Tilmann von Rosen[1,2], MA, MBBS; Peter Tinnemann[1], MPH, MD, MBBS; Falk Müller-Riemenschneider[1,3], MSc, PD, MD, MBBS

[1]Institute for Social Medicine, Epidemiology and Health Economics, Charité – Universitätsmedizin Berlin, Berlin, Germany

[2]Institute for Public Health and Nursing Research, University of Bremen, Bremen, Germany

[3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

Corresponding Author:
Antonella Juline von Rosen, MA, MBBS
Institute for Social Medicine, Epidemiology and Health Economics
Charité – Universitätsmedizin Berlin
Luisenstr 57
Berlin, 10117
Germany
Phone: 49 15208804891
Email: an_pa@uni-bremen.de

Abstract

Background: The Internet is widely used by adolescents for sexual health information and bears the potential to increase knowledge and positively affect behavior.

Objective: The objective of this study is to assess students' preferences when looking for sexual health information online.

Methods: We conducted a cross-sectional survey among ninth grade students in a convenience sample of 13 secondary schools in Berlin, Germany. During a regular school period, participants were requested to rate the importance they attribute to nine aspects of sexual health websites in a paper-based questionnaire. Bivariate and multivariable analyses were used to assess awareness and preferences by gender, age, migrant background, and school type.

Results: Of 1190 eligible students, 1177 (98.91%) students with a mean age of 14.6 (SD 0.7) years participated, 52.52% (605/1152) were male, and 52.94% (612/1156) had at least one parent born abroad. Participant numbers were spread equally across three types of secondary schools in Berlin. Website aspects most frequently cited as important were easily comprehensible wording (88.33%, 961/1088), clear information layout (80.57%, 871/1081), and reliability of the website's publisher (79.28%, 857/1081), whereas the visual style of a website was deemed important by the lowest number of students (35.13%, 378/1076). There was a marked gender difference in the importance students attached to website publisher reliability. Although 437/515 (84.9%) of female participants regarded this as important, only 420/566 (74.2%) of male participants did likewise (P<.001). In multivariable analyses, demographic differences were also particularly visible in the importance of publisher reliability: male participants were significantly less likely to find this aspect important (OR 0.50, 95% CI 0.37-0.69). The odds ratio for students with migrant background was 0.64 (95% CI 0.50-0.81, reference=no migrant background) and OR 2.04 (95% CI 1.03-4.03) for students in the most academic school type (reference=least academic).

Conclusions: Students prefer easily understandable online resources. Setting up sexual health websites according to the explicit preferences of the target audience might encourage usage, especially by those subpopulations less likely to critically assess information validity: male adolescents, children of immigrants, and the academically disadvantaged.

(J Med Internet Res 2017;19(11):e379) doi:10.2196/jmir.7068

KEYWORDS

adolescent; adolescent behavior; Internet; reproductive health; health literacy; sex education; cross-sectional studies; online preferences; eHealth literacy

Introduction

The Internet as a Source of Sexual Health Information

The rapid expansion of the Internet [1,2] has led to an information revolution that affects nearly all aspects of our lives. It completely changed the way we access, analyze, and use information. This holds true even for areas as delicate and private as health. In the United States, searching for health information became the third most popular online activity for all Internet users in 2010 [3]. Several studies show that the vast majority of adolescents look for health information online [4-6]. Of particular interest for teenagers is sexual health, an area raising a multitude of questions perceived as embarrassing, controversial, or sensitive [5-9]. The threshold for adolescents to search for answers to these questions is lowered by the anonymous nature of the Internet, its easy and quick availability, and its low cost [10,11]. A digital divide has been widely described, with socioeconomically disadvantaged subpopulations less likely to have access to the Internet and to use it regularly [12]. However, recent evidence shows that the divide regarding Internet access is closing in industrialized countries, especially among adolescents and young adults. A representative survey in Germany in 2016 showed that across socioeconomic and ethnic divides, 100% of respondents aged 16 to 24 years had Internet access [13], with 97% stating daily usage [14].

The Internet hence bears the potential to increase sexual health knowledge and promote healthy behavior, especially in socioeconomic and ethnic groups with lower levels of access to traditional sources of sexual health information [5,15]. However, reputable sexual health websites with sound content coexist with a vast array of websites presenting incorrect or incomplete information that can misinform and might ultimately lead to unhealthy behavior [16,17]. In the competitive market for online patronage, user centricity has been suggested as an important advantage [18]. Providers of sexual health information, such as public health agencies, should strive to offer online resources that closely match the requirements and preferences of their target population.

To date, evidence on adolescents' health website preferences is scarce and exclusively qualitative. The sole study focusing specifically on adolescents' evaluation of a sexual health website found that interviewees frequently disliked text-heavy style and appreciated interactive features and "real-life stories" from other adolescents [19]. A study on adolescents' evaluation of a general health website yielded similar results [20]. A study on adolescents' website preferences, albeit with a mental health focus, concluded that interviewees were highly critical of website publisher credibility, and that academically disadvantaged participants in particular preferred low text density [21].

To our knowledge, no quantitative study design has to date been employed to assess adolescents' sexual health website preferences. Although qualitative research is clearly useful in gaining an in-depth understanding of preferences, a quantitative study can offer insight into preferences of a wider sample and highlight and quantify differences between different subpopulations.

Study Objective

The objective of this study is to assess which aspects of a website adolescents deem important when they look for sexual health information online, and to assess differences between demographic subpopulations.

Methods

Study Design

Data were collected as part of a larger cross-sectional study on sexual health knowledge that was conducted throughout the year 2012 in the ninth grades of secondary schools in Berlin, Germany. The study, its methodology, and participant demography have been partially described elsewhere [22].

The study design was approved by the Ethics Committee of the Charité – Universitätsmedizin Berlin, as well as the Berlin Senate's Department for Education, Youth and Science. In accordance with legal requirements, the study was discussed and approved by parent-teacher conferences in all schools prior to study onset. As demanded by state law, students aged 13 years and younger were required to provide written parental consent.

Reporting on this study was based on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement guidelines [23].

Sampling and Data Collection

The project study group contacted the heads of the biology or natural science departments of all public secondary schools by phone in late 2011 and supplied schools with further information on the study via email. Of a total of 287 eligible schools, the heads of department were successfully contacted in 142 cases, with 13 schools agreeing to participate. In all 13 schools, the parent-teacher conferences voted in favor of participation. Of the schools choosing not to partake, all quoted time constraints of teaching staff as the reason.

Since a major education reform implemented in 2010, state secondary education in Berlin formally consists of two school forms: university-preparatory schools [Gymnasium] and Integrated Secondary Schools [Integrierte Gesamtschule]. An important subdivision can be made in the latter category between schools allowing for their highest-achieving students to continue up to year 12 to qualify for university access, and schools where no such option exists and most students are likely not to continue their formal education past year 10. For brevity, these three school types were reported in results as "highest (academic) tier" (gymnasium), "intermediate (academic) tier" (integrated secondary school with option to qualify for university), and "lowest (academic) tier" (integrated secondary school up to year 10 only). All three school types were represented in the sample, as were seven of the 12 Berlin city districts (Mitte, Pankow, Charlottenburg-Wilmersdorf, Spandau, Steglitz-Zehlendorf, Treptow-Köpenick, and Marzahn-Hellersdorf). Districts from both the former Eastern and Western parts of Berlin were included, and districts with participating schools ranged from inner city to suburban and from the most affluent to relatively deprived [24].

Study Setting

The survey was conducted by a member of the project team during a school period in the regular classroom setting. A short presentation informed students on the aim of the study and of the voluntary and anonymous nature of participation. It was emphasized that results would not be disclosed to teaching staff, parents, or fellow students. To safeguard anonymity, it was stressed that it was strictly prohibited to attempt to read other students' responses, and an exam-like, nonclustered seating layout in the classroom was adopted. Subsequently, a paper-based questionnaire was administered to all participating students. A member of the research team supervised the class until all questionnaires had been collected.

Questionnaire

Due to the lack of existing tools for the research question, the study questionnaire was devised by the authors to assess students' preferences and requirements when looking for sexual health information online. To ascertain the adequacy of the questionnaire for the adolescent sample population, a pretest was performed in one school class, with comments from students leading to minor modifications. Following discussion with students from the pretest sample, wording was phrased indirectly ("How important would you find the following aspects when looking for sexual health information on the Internet?"). Importance was rated on a five-point Likert scale, on which students were to choose between "important," "somewhat important," "neither important nor unimportant," "rather unimportant," and "unimportant" (authors' translation from German).

Demographic variables on gender, age, and migrant background (parental place of birth) were included in the questionnaire. For the latter, response options were restricted to "Germany" or "other country" in accordance with the principle of parental informational self-determination.

Aspects were chosen by the authors following an in-depth analysis of the different characteristics and features of 10 prominent German-language websites providing sexual health information: Loveline [25], Mach's mit [26], profamilia [27], Sexundso [28], sextra [29], BRAVO [30], gutefrage [31], mädchen.de [32], gofeminin [33], and Lovetalk [34]. Websites were chosen based on a search on Google for the following keywords and combinations thereof: adolescents, students, love, sex, first time, contraception, condom, pill, pregnancy, STIs, and HIV (authors' translation of the German keywords). Google was used as the most frequently cited starting point for sexual health information online in previous research [35]. A total of nine aspects belonging to three principal groups were identified: "who and for whom" (publisher and explicit target audience of a website), "information presentation" (visual style, clear layout of information, easily understandable language, and text shortness), and "website features" (facilities to ask individual questions, a section where people can report their own personal experiences, and advice provided by people of similar age). The websites as well as the nine website aspects were discussed with and deemed appropriate by our pretest sample.

Statistics and Data Analysis

For the analysis of the survey data, IBM SPSS version 23 was used. Descriptives were computed. For the statistical analyses, outcomes were dichotomized for the sake of readability: the variables "important" and "rather important" were clustered to "(rather) important," as were the categories "of medium importance," "rather unimportant," and "unimportant" to "other response." Descriptives and chi-square tests for the original five-point outcome scale are included in Multimedia Appendix 1.

Descriptives and Bivariate Analysis

Aspects were ranked by the aggregate percentage of students selecting "(rather) important." Using two-sided chi-square tests, bivariate relationships were calculated between the demographic variables age, gender, migratory background, and school type and the outcome variables measuring the importance participants attribute to the different aspects of sexual health information websites. For statistical analysis, migrant background was defined as having at least one parent who was born abroad.

Multivariable Analysis

Regression models were used to quantify the effect of age, gender, migratory background, and school type on outcome variables. Due to the clustered nature of observation in class and school groups, a mixed multilevel regression model (SPSS GENLINMIXED) containing school and class as random effects was used. Odds ratios and 95% confidence intervals for all outcome variables were computed from mixed multilevel regression. To account for possible violations of model assumptions, we used robust estimation. Missing cases were excluded from bivariate and regression analysis.

Results

Population

The study was conducted in 61 school classes. A total of 1190 students were present in class on the day of the study and thus eligible for participation. Ten students aged 13 years could not participate due to missing parental consent and a further two students chose not to participate. One student could not participate due to lacking the required basic language proficiency. Therefore, a total of 1177 participants were included in the survey, equating to a response rate of 98.91%. The overall mean age of participants was 14.6 (SD 0.8) years, the mean for boys being 14.6 (SD 0.7) years and 14.5 (SD 0.7) years for girls. Table 1 shows the demography of study participants, as reported previously [22].

Descriptives and Bivariate Analysis

Table 2 shows the distribution of responses for the items in the three groups of website aspects, the relative rank of each aspect by cumulative percentage selecting "(rather) important," and stratification by gender.

Table 1. Demography of participants (N=1177).

Population characteristic	n (%)
Age (years)	**n=1162[a]**
13	23 (1.98)
14	565 (48.62)
15	480 (41.31)
16	94 (8.09)
Gender	**n=1152[a]**
Female	547 (47.48)
Male	605 (52.52)
Migrant background	**n=1156[a]**
Both parents born in Germany	544 (47.06)
Both parents born abroad	352 (30.45)
Only mother born abroad	127 (10.99)
Only father born abroad	133 (11.51)
School type	**n=1177[b]**
Lowest tier	390 (33.14)
Intermediate tier	395 (33.56)
Highest tier	392 (33.31)

[a]Total number of responses for each variable.

[b]Entered for all participants at school level.

Table 2. Website preferences by gender (bivariate analysis).

Website aspect group and website aspect	Aspect rank[a]	Population in (rather) important, n (%)			*P* [b]
		Female	Male	Total	
Who and for whom					
Reputable publisher (n=1081[c])	3	437 (84.85)	420 (74.20)	857 (79.28)	<.001
Explicitly addressed at adolescents (n=1088[c])	7	366 (70.66)	311 (54.56)	677 (62.22)	<.001
Information presentation					
Language easily understandable (n=1088[c])	1	463 (89. 38)	498 (87.37)	961 (88.33)	.35
Information clearly laid out (n=1081[c])	2	425 (83.17)	446 (78.25)	871 (80.57)	.045
Texts short and concise (n=1084[c])	8	260 (50.58)	304 (53.33)	564 (52.03)	.39
Visual style/design attractive (n=1076[c])	9	149 (29.22)	229 (40.46)	378 (35.13)	<.001
Website features					
Possibility to ask questions (n=1079[c])	4	378 (74.14)	406 (71. 35)	784 (72.66)	.34
Section with personal experiences (n=1080[c])	5	361 (70.23)	381 (67.31)	742 (68.70)	.32
Advice by other adolescents (n=1072[c])	6	336 (66.14)	373 (66.13)	709 (66.14)	>.99

[a]By percentage (rather) important.

[b]Calculated from chi-square tests.

[c]Number of participants included in the analysis.

Student Preferences: Who and for Whom

Publisher reliability was described as (rather) important by 857 of 1081 students (79.28%), the percentage for girls being higher than for boys. Students of migrant background were significantly less likely to attach importance to this aspect (*P*<.001), as were students in the lowest academic tier of schools (*P*=.01). Whether a website is explicitly addressed at adolescents was an aspect to which students attributed relatively little importance. Again, the aspect was considerably more important to girls, although there was no association with the other demographic variables.

Student Preferences: Information Presentation

Understandable language was the most important aspect overall. Importance was higher in the highest academic school tier (*P*=.001) and lower for children of immigrants (*P*<.001). Across the sample, 871 of 1081 respondents (80.57%) described it as (rather) important that information should be clearly laid out. Girls were significantly more likely than boys to regard this aspect as (rather) important. Again, school type and migratory background were significantly associated with outcomes: the higher the school tier, the more important the issue was considered (*P*=.03), and children of immigrants were less likely to regard the aspect as (rather) important (*P*<.001).

Whether information text on sexual health websites was short and concise was an issue of relatively low importance in our survey, with 564 of 1084 students (52.03%) attaching importance. Students from the lowest tier of schools and children of immigrants were significantly more likely to regard this aspect as (rather) important (*P*=.01 for both).

Visual style was the least important issue overall. Although no significant difference by migratory background was observed, school type was significantly associated with the importance attached to a website's visual style: students from the lowest tier of schools were significantly more likely to regard this aspect as (rather) important (*P*=.01).

Student Preferences: Website Features

Among participating students, 784 of 1079 (72.66%) found it (rather) important that sexual health websites offer the possibility to ask individual questions, thus putting it in fourth place overall. It was followed in fifth place by the aspect that websites should include a section where people can report their own personal experiences and in sixth place by whether websites included advice written by people of similar age as the user. Students with migrant background were less likely to deem personal experiences and writing by age peers (rather) important (*P*=.04 and *P*=.03, respectively). Otherwise, for website feature variables, no significant difference by gender, age, migratory background, or school type was observed.

Multivariable Analysis

Results from the regression model are presented in Table 3. Differences in odds were especially marked on the issue of the reputability of a website's publisher: female students were nearly twice as likely to attach importance to this aspect, with not having a migrant background and attending a school in the highest academic tier also greatly increasing the odds to select this option. Being of female gender also increased the odds of attributing importance to whether a website was explicitly addressed at adolescents by more than twofold.

Table 3. The effect of the demographic variables on outcomes (multivariable analysis).

Aspect group and website aspect	Variable, OR (95% CI)[a]				
	Age (per year increase)	Gender (male)	Migration (migrant background)	School type (intermediate tier)	School type (highest tier)
Who and for whom					
Reputable publisher (n=1074[b])	1.23 (1.03-1.47)[c]	0.50 (0.37-0.69)[c]	0.64 (0.50-0.81)[c]	1.39 (0.70-2.73)	2.04 (1.03-4.03)[c]
Explicitly addressed at adolescents (n=1081[b])	1.16 (0.96-1.40)	0.48 (0.35-0.67)[c]	0.76 (0.66-0.88)[c]	1.08 (0.74-1.57)	1.02 (0.83-1.26)
Information presentation					
Language easily understandable (n=1081[b])	1.02 (0.91-1.13)	0.85 (0.51-1.39)	0.47 (0.29-0.75)[c]	1.33 (0.70-2.53)	2.62 (1.36-5.06)[c]
Information clearly laid out (n=1074[b])	1.13 (1.01-1.27)[c]	0.71 (0.48-1.06)	0.71 (0.48-1.04)	1.10 (0.42-2.86)	1.75 (0.85-3.59)
Texts short and concise (n=1077[b])	1.32 (1.10-1.58)[c]	1.02 (0.84-1.25)	1.46 (1.10-1.94)[c]	0.71 (0.50-1.03)	0.81 (0.46-1.44)
Visual style/design attractive (n=1069[b])	1.15 (1.00-1.32)[c]	1.57 (1.08-2.29)[c]	1.02 (0.78-1.32)	0.68 (0.52-0.90)[c]	1.07 (0.70-1.52)
Website features					
Possibility to ask questions (n=1072[b])	1.00 (0.82-1.24)	0.87 (0.71-1.07)	0.89 (0.64-1.24)	1.06 (0.71-1.58)	1.14 (0.80-1.62)
Section with personal experiences (n=1073[b])	1.17 (1.02-1.33)[c]	0.84 (0.63-1.14)	0.88 (0.69-1.11)	0.87 (0.48-1.60)	1.17 (0.68-2.01)
Advice from other adolescents (n=1065[b])	1.08 (0.06-8.62)	0.99 (0.76-1.23)	0.78 (0.59-1.04)	0.91 (0.61-1.37)	1.27 (0.92-1.75)

[a]Reference category for gender was female, migration was none, and for both school types was lowest tier.

[b]Number of participants included in the analysis.

[c]*P*<.05.

Discussion

We conducted a study among ninth graders in Berlin secondary schools to evaluate adolescents' preferences when looking for sexual health information online. Despite the voluntary nature of participation, only two of 1179 students chose not to participate. The overwhelming rate of participation might be an expression of a strong interest and curiosity invoked by the topic of sexual health and Internet research among adolescents. To our knowledge, no quantitative study with a similar focus on adolescent preferences in online resources has been conducted to date.

Easily comprehensible wording was most frequently selected as (rather) important. This was followed by clear information layout. In third place came the credibility of a website's publisher, which was described as (rather) important by about 80% of participants. We found that male gender, migratory background, and attending a school of the lowest academic tier were significantly associated with lower importance placed on publisher reliability in both bivariate and multivariable analyses. The possibility to ask questions, whether websites contained a section with personal experiences, and whether websites included advice from other adolescents were of intermediate relative importance to students in our sample. Interestingly, results for website feature aspects were relatively uniform across demographic divides. Both the visual style and the youth specificity of websites were relatively unimportant to study participants. Male respondents were more likely to regard this aspect as (rather) important.

The Internet has enabled direct access to a vast array of health information, which was previously only available through intermediates such as health care professionals. The term "disintermediation" has been coined to describe this development [36]. It promises greater access to health information, especially in sensitive areas such as sexual health [9]. However, it also bears the risk that incorrect information is accessed and perceived as accurate [17,37]. This is emphasized by the visible minority of students in our sample neglecting publisher credibility of sexual heath websites. From a public health perspective, a dual strategy should be employed to facilitate the access of adolescents to reliable sexual health information online. Firstly, health care professionals and educators can provide a certain degree of apomediation. Secondly, access of reliable and suitable sexual health information resources should be encouraged by adapting websites to explicitly cater to the target audience's preferences, and by educating adolescents to critically access the validity of sexual health information online.

Apomediation describes the process in which individuals or institutions provide guidance in the digital sphere, for example by recommending particular websites [36]. One study has suggested that health care professionals can play a pivotal role in providing a list of suitable and reliable health websites as a first port of call [38]. The reach of such an intervention among adolescents might be even higher in Germany than in other contexts given that statutory health insurance covers and promotes a one-time routine checkup visit to a physician for all youths aged 12 to 14 years [39]. Thus, physicians can extend their reach to healthy teenagers, and access is not impeded by financial considerations for adolescents or parents. A similar role with even greater reach could be played by schools. Peer apomediation, such as other adolescents recommending sexual health resources through social media, also bears the potential to spread knowledge regarding sexual health resources online. Further studies are needed to evaluate the reach and effectiveness of professional and peer apomediation.

Although recommendations of trustworthy websites by health care professionals, teachers, and possibly peers could play a role in providing adolescents with websites to start their search for sexual health information, it has to be taken into account that many users access health information including sexual health information starting with search engine queries rather than through apomediation [35,40,41]. User centricity is a key factor in preventing consumers from quickly "bouncing" back to search results [18]. Especially in the field of health where incorrect or misleading information can undermine healthy behavior, it is imperative that websites by reputable providers of sexual health information are set up according to the preferences and requirements of users. Our study can contribute to an understanding of adolescent preferences in sexual health websites. The most frequently cited aspect in our study, easily comprehensible language, should be actively pursued in the development of all sexual health resources. Beyond matching their preferences, it is axiomatic that information that users can linguistically comprehend is more likely to improve knowledge and potentially encourage healthy behavior. However, studies show that many health websites fall short in this regard and employ a language prohibitively sophisticated for many users, especially the young and/or educationally disadvantaged [42,43]. Our results likely reflect similar experiences of participants' with overly complicated sexual health information online. Sexual health websites should hence employ accessible language as a dually user-centric measure: to make a website both more used by and more useful to adolescents.

One way to further user centricity is to invite target group participation at the different steps of development [19]. Because our study shows significant divergence between preferences even within a narrow age bracket, it is critical that participation is invited from adolescents across genders and ethnic and educational backgrounds. It can be hypothesized that reputable sexual health information providers, such as public health authorities, government bodies, or sexual health associations, could use this as a competitive advantage: their bona fide standing is likely to make it easier to co-opt youths to participate in website development, such as through schools.

In some cases, differences between demographic groups might warrant the setting up of separate resources. For example, our study shows pronounced differences in preferences between male and female adolescents. One solution would be to develop sexual health websites aimed at one gender only, or to set up male and female subsites to be selected by users within one sexual health website. Indeed, a previous study has shown that a majority of adolescents would prefer gender-stratified sexual health websites [20].

However, there are many cases in which it will not be possible to develop separate resources for subpopulations with divergent preferences. One innovative way to make online health resources palatable to a diverse target audience is to enable users to interactively modify different parameters of a website [44]. For example, users could select between a text-heavy and a more visual presentation of information or between elaborate and short articles. This enables users to adapt the website not only to general personal preferences, but also according to the specific search context and purpose. This innovative and user-centric concept has been implemented and positively evaluated for a general health information database [44].

Lastly, it is important that adolescents are educated and encouraged to access, comprehend, and critically evaluate online health information and sexual health information. The term "eHealth literacy" has been coined for this set of skills [45-47]. An integral part of eHealth literacy is the ability to appraise the validity of online information [48]. In our study, approximately 80% of students described the reliability of a website's publisher as (rather) important. This is consistent with previous studies that found the majority of users are critical when looking for health information online, whereas a visible minority tend to neglect the issue of credibility of health websites [49-51]. We found that male gender, migratory background, and attending a school of the lowest academic tier were associated with lower importance placed on publisher reliability. With strong evidence that these groups are also more likely to have poor sexual health knowledge and to engage in risk-taking behavior both in Germany [15] and in other settings [52,53], it is crucial to educate students to critically evaluate the reliability of information resources, especially in these vulnerable subpopulations. Particularly in countries with mandatory school education, such as Germany, schools could be instrumental in providing health and sexual health online literacy education to virtually all adolescents. For this, the significant institutional inertia of school policy has to be overcome and joint initiatives at the different levels of school policy are called for [54].

A key strength of this study is the large study population including students from all types of public schools in Berlin. Furthermore, a voluntary participation rate of 98.9% of students present in class makes it unlikely that participation/nonparticipation introduced a significant bias.

Unlike the student level, in which the rate of participation was very high, it has to be acknowledged that less than 5% of Berlin secondary schools participated. Although a majority of Berlin's city districts and very diverse academic, socioeconomic, and geographic settings were represented in the sample, the generalizability of results might be limited by a systematic difference between participating and nonparticipating schools. Furthermore, it should be considered that findings of adolescents in Berlin might not be generalizable to other geographic settings.

A further limitation of this study is the fact that students were asked to evaluate the importance of different website aspects when looking for sexual health information online, without reference to specific websites during the survey. The abstract thinking this requires might have been challenging for at least some of the students in the sample. Future studies could include a more concrete evaluation of website aspects. For example, qualitative methods such as focus group interviews could be employed to have students evaluate different sexual health websites or to gain a more in-depth understanding of by what means and how accurately students evaluate publisher reliability. Furthermore, website analytics could be employed to quantify how successfully different sexual health website formats are in attracting and retaining visitors.

Websites providing sexual health information for adolescents bear the potential to improve sexual health knowledge and promote healthy behavior. To be effective, websites should be set up according to the preferences and requirements of the target population not merely regarding content, but also regarding the way information is presented. Divergent preferences might warrant the establishment of websites specifically geared at specific subpopulations, such as male adolescents or adolescents with a migrant background. Furthermore, parents, schools, and public health authorities should strive to improve online literacy among adolescents to make sure they have the facilities to critically evaluate the reliability of information online—in all aspects of information access, but especially in the critical field of sexual health.

Acknowledgments

This study was conducted within the framework of a research project to improve the sexual health knowledge of adolescents in Germany. The project was partially funded by the German Federal Ministry of Education and Research. The sponsor did not influence the design of the study; the collection, analyses, or interpretation of data; the writing of the manuscript; or the decision to publish the results.

The authors express their gratitude to all schools and students participating in the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Unclustered Likert-Scale outcomes of website preferences by gender.

[PDF File (Adobe PDF File), 31KB - jmir_v19i11e379_app1.pdf]

References

1. Statistisches Bundesamt. 2016. Staat & Gesellschaft - Ausstattung mit Gebrauchsgütern URL: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/AusstattungGebrauchsguetern/Tabellen/ZeitvergleichAusstattung_IKT.html;jsessionid=722C21068A1FCFBEE9C222F124BC6AB9.cae4 [accessed 2016-08-07] [WebCite Cache ID 6jaePHtDc]
2. United States Census Bureau. 2012. Computer and Internet use in the United States: 2010 URL: http://www.census.gov/prod/2013pubs/p20-569.pdf [accessed 2015-03-26] [WebCite Cache ID 6XK03zZGC]
3. Zickuhr K. Pew Research Center. 2010 Dec 16. Generations 2010 URL: http://www.pewinternet.org/2010/12/16/generations-2010/ [accessed 2016-08-07] [WebCite Cache ID 6jafBXNuO]
4. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
5. Borzekowski DL, Rickert VI. Adolescent cybersurfing for health information: a new resource that crosses barriers. Arch Pediatr Adolesc Med 2001 Jul;155(7):813-817. [Medline: 11434849]
6. Skinner H, Biscope S, Poland B, Goldberg E. How adolescents use technology for health information: implications for health professionals from focus group studies. J Med Internet Res 2003 Dec 18;5(4):e32 [FREE Full text] [doi: 10.2196/jmir.5.4.e32] [Medline: 14713660]
7. Ralph LJ, Berglas NF, Schwartz SL, Brindis CD. Finding teens in TheirSpace: using social networking sites to connect youth to sexual health services. Sex Res Soc Policy 2011 Feb 22;8(1):38-49. [doi: 10.1007/s13178-011-0043-4]
8. Selkie EM, Benson M, Moreno M. Adolescents' views regarding uses of social networking websites and text messaging for adolescent sexual health education. Am J Health Educ 2011 Dec;42(4):205-212 [FREE Full text] [Medline: 22229150]
9. Racey M, Machmueller D, Field D, Kulak V, Newton GS. Perceptions and use of sources of health knowledge by young adolescents. Int J Adolesc Med Health 2016 Jun 14:pii. [doi: 10.1515/ijamh-2016-0002] [Medline: 27299195]
10. Suzuki LK, Calzo JP. The search for peer advice in cyberspace: An examination of online teen bulletin boards about health and sexuality. J Appl Dev Psychol 2004 Nov;25(6):685-698. [doi: 10.1016/j.appdev.2004.09.002]
11. Henderson EM, Keogh E, Eccleston C. Why go online when you have pain? A qualitative analysis of teenagers' use of the Internet for pain management advice. Child Care Health Dev 2014 Jul;40(4):572-579. [doi: 10.1111/cch.12072] [Medline: 23663232]
12. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
13. Statistisches Bundesamt. 2016. Staat & Gesellschaft - IT-Nutzung URL: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/ITNutzung/ITNutzung.html [WebCite Cache ID 6ovXXPb05]
14. Statistisches Bundesamt. 2016. Private Nutzung von Informations- und Kommunikationstechnologien 2016: Durchschnittliche Nutzung des Internets URL: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/ITNutzung/Tabellen/NutzungInternetAlter_IKT.html [accessed 2017-03-13] [WebCite Cache ID 6ovXGSihG]
15. Heßling A, editor. Youth Sexuality 2010: Repeat Survey of 14 to 17-year-olds and their parents. Cologne: Bundeszentrale für Gesundheitliche Aufklärung (BZgA); 2010.
16. Kanuga M, Rosenfeld WD. Adolescent sexuality and the internet: the good, the bad, and the URL. J Pediatr Adolesc Gynecol 2004 Apr;17(2):117-124. [doi: 10.1016/j.jpag.2004.01.015] [Medline: 15050988]
17. Madden T, Cortez S, Kuzemchak M, Kaphingst KA, Politi MC. Accuracy of information about the intrauterine device on the Internet. Am J Obstet Gynecol 2016 Apr;214(4):499.e1-499.e6. [doi: 10.1016/j.ajog.2015.10.928] [Medline: 26546848]
18. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
19. McCarthy O, Carswell K, Murray E, Free C, Stevenson F, Bailey JV. What young people want from a sexual health website: design and development of Sexunzipped. J Med Internet Res 2012;14(5):e127 [FREE Full text] [doi: 10.2196/jmir.2116] [Medline: 23060424]
20. Franck LS, Noble G. Here's an idea: ask the users! Young people's views on navigation, design and content of a health information website. J Child Health Care 2007 Dec;11(4):287-297. [doi: 10.1177/1367493507083941] [Medline: 18039731]
21. Havas J, de Nooijer J, Crutzen R, Feron F. Adolescents' views about an internet platform for adolescents with mental health problems. Health Educ 2011 Apr 19;111(3):164-176. [doi: 10.1108/09654281111123466]
22. von Rosen FT, von Rosen AJ, Müller-Riemenschneider F, Tinnemann P. Awareness and knowledge regarding emergency contraception in Berlin adolescents. Eur J Contracept Reprod Health Care 2017 Feb;22(1):45-52. [doi: 10.1080/13625187.2016.1269162] [Medline: 28056564]
23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007 Oct 20;370(9596):1453-1457. [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]

24. Amt für Statistik Berlin-Brandenburg. 2016. Regionaler Sozialbericht Berlin und Brandenburg 2015 URL: https://www.statistik-berlin-brandenburg.de/produkte/pdf/SP_Sozialbericht-000-000_DE_2015_BBB.pdf [accessed 2017-04-02] [WebCite Cache ID 6pQ5WHLoe]
25. Loveline - das Jugendportal der Bundeszentrale für gesundheitliche Aufklärung BZgA. URL: https://www.loveline.de/startseite.html [accessed 2017-10-25] [WebCite Cache ID 6uTPpJBLp]
26. Mach's mit. URL: https://www.liebesleben.de/fuer-alle/ [accessed 2017-10-25] [WebCite Cache ID 6uTPvEe4Q]
27. profamilia. URL: https://www.profamilia.de// [accessed 2017-10-25] [WebCite Cache ID 6uTQK4GDi]
28. Sexundso – Online Beratung. URL: http://www.sexundso.de [accessed 2017-10-25] [WebCite Cache ID 6uTQKjLlf]
29. sextra. URL: https://profamilia.sextra.de/ [accessed 2017-10-25] [WebCite Cache ID 6uTQLTPxr]
30. BRAVO. URL: http://www.bravo.de/ [accessed 2017-10-25] [WebCite Cache ID 6uTQbuMsy]
31. gutefrage. URL: https://www.gutefrage.net/ [accessed 2017-10-25] [WebCite Cache ID 6uTQRelJY]
32. Mädchen.de. URL: http://www.maedchen.de [accessed 2017-10-25] [WebCite Cache ID 6uTQSBddA]
33. gofeminin. URL: http://www.gofeminin.de [accessed 2017-10-25] [WebCite Cache ID 6uTQWeaQp]
34. Lovetalk – das Forum für Liebe und mehr. URL: https://www.lovetalk.de/ [accessed 2017-10-25] [WebCite Cache ID 6uTQX8Nhg]
35. Buhi ER, Daley EM, Fuhrmann HJ, Smith SA. An observational study of how young people search for online sexual health information. J Am Coll Health 2009;58(2):101-111. [doi: 10.1080/07448480903221236] [Medline: 19892646]
36. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
37. Kortum P, Edwards C, Richards-Kortum R. The impact of inaccurate Internet health information in a secondary school learning environment. J Med Internet Res 2008 Jun 30;10(2):e17 [FREE Full text] [doi: 10.2196/jmir.986] [Medline: 18653441]
38. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
39. Traub U, Schmierer R, Kauth T, Schönauer T. J1-Initiative: "Every hero needs a doctor"; Neuere Ansätze zur Verbesserung der Teilnahme an der Jugendgesundheitsuntersuchung (J1). Gesundheitswesen 2015 Apr 9;77(04):1. [doi: 10.1055/s-0035-1546902]
40. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
41. Pang PC, Verspoor K, Chang S, Pearce J. Conceptualising health information seeking behaviours and exploratory search: result of a qualitative study. Health Technol 2015 Feb 10;5(1):45-55. [doi: 10.1007/s12553-015-0096-0]
42. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA 2001;285(20):2612-2621 [FREE Full text] [Medline: 11368735]
43. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
44. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
45. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of "Dr Google": a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
46. Agree EM, King AC, Castro CM, Wiley A, Borzekowski DLG. "It's got to be on this page": age and cognitive style in a study of online health information seeking. J Med Internet Res 2015;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]
47. 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]
48. 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]
49. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. The Internet: a window on adolescent health literacy. J Adolesc Health 2005 Sep;37(3):243. [doi: 10.1016/j.jadohealth.2004.08.023] [Medline: 16109345]
50. Jones RK, Biddlecom AE. Is the internet filling the sexual health information gap for teens? An exploratory study. J Health Commun 2011 Feb;16(2):112-123. [doi: 10.1080/10810730.2010.535112] [Medline: 21207311]
51. Hargittai E, Young H. Searching for a "Plan B": young adults' strategies for finding information about emergency contraception online. Policy Internet 2012 Mar;4(2):1-23. [doi: 10.1515/1944-2866.1176]
52. Santelli JS, Lowry R, Brener ND, Robin L. The association of sexual behaviors with socioeconomic status, family structure, and race/ethnicity among US adolescents. Am J Public Health 2000 Oct;90(10):1582-1588. [Medline: 11029992]

53. de Graaf H, Vanwesenbeeck I, Meijer S. Educational differences in adolescents' sexual health: a pervasive phenomenon in a national Dutch sample. J Sex Res 2015;52(7):747-757. [doi: 10.1080/00224499.2014.945111] [Medline: 25260077]
54. Harris FJ. Challenges to teaching credibility assessment in contemporary schooling. In: Metzger MJ, Flanagin AJ, editors. Digital Media, Youth, and Credibility. Cambridge, MA: MIT Press; 2008:155-180.

Edited by M Focsa; submitted 29.11.16; peer-reviewed by S Chang, D Levine, J Bailey; comments to author 21.12.16; revised version received 03.04.17; accepted 20.09.17; published 08.11.17.

Please cite as:
von Rosen AJ, von Rosen FT, Tinnemann P, Müller-Riemenschneider F
Sexual Health and the Internet: Cross-Sectional Study of Online Preferences Among Adolescents
J Med Internet Res 2017;19(11):e379
URL: http://www.jmir.org/2017/11/e379/
doi:10.2196/jmir.7068
PMID:29117927

Original Paper

Reducing Symptom Distress in Patients With Advanced Cancer Using an e-Alert System for Caregivers: Pooled Analysis of Two Randomized Clinical Trials

David H Gustafson[1], PhD; Lori L DuBenske[2], PhD; Amy K Atwood[1], PhD; Ming-Yuan Chih[3], PhD; Roberta A Johnson[1,4], MA, MAEd; Fiona McTavish[1], MS; Andrew Quanbeck[1,4], PhD; Roger L Brown[5], PhD; James F Cleary[6], MD; Dhavan Shah[7], PhD

[1]Department of Industrial and Systems Engineering, Center for Health Enhancement Systems Studies, University of Wisconsin-Madison, Madison, WI, United States

[2]Department of Psychiatry, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States

[3]Department of Clinical Sciences, College of Health Sciences, University of Kentucky, Lexington, KY, United States

[4]Department of Family Medicine and Community Health, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States

[5]Nursing Research Department, School of Nursing, University of Wisconsin-Madison, Madison, WI, United States

[6]Medical Oncology Section, Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States

[7]Mass Communication Research Center, School of Journalism and Mass Communication, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
David H Gustafson, PhD
Center for Health Enhancement Systems Studies
Department of Industrial and Systems Engineering
University of Wisconsin-Madison
Mechanical Engineering Building, 4th Floor
1513 University Ave.
Madison, WI, 53706
United States
Phone: 1 6082395535
Fax: 1 6088901438
Email: dhgustaf@wisc.edu

Abstract

Background: Symptom distress in patients toward the end of life can change rapidly. Family caregivers have the potential to help patients manage those symptoms, as well as their own stress, if they are equipped with the proper resources. Electronic health (eHealth) systems may be able to provide those resources. Very sick patients may not be able to use such systems themselves to report their symptoms but family caregivers could.

Objective: The aim of this paper was to assess the effects on cancer patient symptom distress of an eHealth system that alerts clinicians to significant changes in the patient's symptoms, as reported by a family caregiver.

Methods: A pooled analysis from two randomized clinical trials (NCT00214162 and NCT00365963) compared outcomes at 12 months for two unblinded groups: a control group (Comprehensive Health Enhancement Support System [CHESS]-Only) that gave caregivers access to CHESS, an online support system, and an experimental group (CHESS+CR [Clinician Report]), which also had CHESS but with a CR that automatically alerted clinicians if symptoms exceeded a predetermined threshold of severity. Participants were dyads (n=235) of patients with advanced lung, breast, or prostate cancer and their respective family caregivers from 5 oncology clinics in the United States of America. The proportion of improved patient threshold symptoms was compared between groups using area-under-the-curve analysis and binomial proportion tests. The proportion of threshold symptoms out of all reported symptoms was also examined.

Results: When severe caregiver-reported symptoms were shared with clinicians, the symptoms were more likely to be subsequently reported as improved than when the symptoms were not shared with clinicians (P<.001). Fewer symptom reports were completed

in the group of caregivers whose reports went to clinicians than in the CHESS-Only group (P<.001), perhaps because caregivers, knowing their reports might be sent to a doctor, feared they might be bothering the clinician.

Conclusions: This study suggests that an eHealth system designed for caregivers that alerts clinicians to worrisome changes in patient health status may lead to reduced patient distress.

Trial Registration: Clinicaltrials.gov NCT00214162; https://clinicaltrials.gov/ct2/show/NCT00214162 (Archived by WebCite at http://www.webcitation.org/6nmgdGfuD) and Clinicaltrials.gov NCT00365963; https://clinicaltrials.gov/ct2/show/NCT00365963 (Archived by WebCite at http://www.webcitation.org/6nmh0U8VP)

(J Med Internet Res 2017;19(11):e354) doi:10.2196/jmir.7466

KEYWORDS

Internet; health communication; palliative care; communication barriers; signs and symptoms; eHealth

Introduction

Managing Patient Symptom Distress

As advanced cancer treatments enable patients to live longer, managing patient symptoms becomes even more important for patients, informal (family or friend) caregivers, and clinicians [1-5]. In some cases, the side effects of advanced treatment (eg, pain and cognitive limitations) can create problems that challenge the fabric of the family [6,7] and sometimes even lead to conflict [6,7] between the family and the clinical team [8]. Some of the problem, from a patient and caregiver perspective, revolves around when and how to get the attention of the clinical team without “bothering” them [9].

Research has shown the importance of timely alerts to providers regarding disconcerting changes in patient conditions. For instance, a study of wait times in the Veterans Health Administration system found that mortality rates increased significantly when patients waited longer to be seen [10]. However, often changes in a patient’s symptoms are reported only at a clinical visit and by then the patient’s condition is often harder to manage than when the symptoms first worsened.

As early as 1997, family caregivers were estimated to provide about US $200 billion of unpaid health care services [11]. Although the burden on family caregivers has been well described [1], the potential of caregivers to improve their lives, and patients’ and providers’ lives, has had comparatively little attention. When properly engaged, the family caregiver can be a critical source of information and support for both the patient and the clinical team. Family caregivers often spend far more time with the patient than anyone else. They can support the patient while collecting information that could be vitally important to providers and for effective care. Some families assume these roles well with no help, whereas others need support to realize their potential. The system studied here was designed to help caregivers maximize their value as a partner in care and minimize their burden.

Recent developments in electronic symptom collection systems [12,13] offer promise for more timely and accurate information, greater patient acceptance, and reduced cost compared with paper-based systems [14]. Studies of such systems have shown moderate to significant improvement of patient symptoms and quality of life [15-19], and even survival [20]. A key issue is when and how to reach clinicians effectively, given how busy they are.

This problem of reporting symptom changes to clinicians is exacerbated for advanced cancer patients. These patients find it increasingly difficult or impossible to use symptom-reporting systems as their disease advances. Informal caregivers are in a position to accurately observe and report on patient symptoms [21-23].

The CHESS System With Reports to Clinicians

The Comprehensive Health Enhancement Support System (CHESS) refers to extensively tested information and communication technologies for coping with cancer and other serious illnesses. The vast majority of CHESS systems have been designed for patients [24-27]. The CHESS "Coping with Cancer" website [28] was designed instead for caregivers to (1) provide them with well-organized cancer, caregiving, and bereavement *information*, (2) serve as a channel to *communicate* with and receive *support* from other caregivers, experts, clinicians, and their social networks, (3) act as a *coach* by gathering information from caregivers and providing feedback based on algorithms (decision rules), and (4) provide *tools* (eg, a program to organize support from family and friends) to improve the caregiving experience. Hence the content and focus of this caregiver program differed substantially from other CHESS systems. The relationship between theory and CHESS was previously reported [29].

The CHESS caregiver system studied in this analysis contained a symptom-reporting system, the Clinician Report (CR), which delivered to the clinical team information about worrisome changes in symptoms collected from informal caregivers of advanced-stage cancer patients [29-31]. Specifically, the CR contacted the clinical team whenever a threshold symptom was reported—that is, when a caregiver rated at least one of 10 patient symptoms at ≥7 on a 0 to 10 severity scale. The alert was intended to quickly bring clinician attention to severe symptoms, potentially leading to timelier symptom management.

Purpose and Contribution of the Study

The study reported here is, to our knowledge, the first to report the effects on patients of an electronic system that collects and analyzes caregiver observations and sends alerts to clinicians when caregivers report worrisome increases in symptom distress in patients suffering from advanced cancer. A previous paper reported the effects on caregivers themselves of using the system [30]. In that study, caregivers with CHESS+CR had less negative mood at both 6 and 12 months than caregivers in the

CHESS-Only group. The groups did not differ significantly on caregiver preparedness or physical burden at either time point. This paper reports the effects of the system on patients, specifically on patient symptom distress. We examined two outcomes: (1) the proportion of improved caregiver-reported, severe ("threshold") symptoms that patients had out of all threshold symptoms and (2) the proportion of caregiver-reported threshold symptoms that patients had out of the total symptoms reported on. Specifically, the study addresses this question: Does the CHESS system with the CR reduce symptom distress in patients more than CHESS without the CR?

Methods

Participants

Between September 2004 and April 2007, 235 dyads of patients with advanced-stage cancer and their primary informal caregivers were recruited to one of two randomized clinical trials of CHESS. One of the trials recruited breast and prostate cancer patients and their caregivers (NCT00214162); the other enrolled lung cancer patients and caregivers (NCT00365963). Eligible breast cancer patients were women with recurrent or metastatic breast cancer. Eligible prostate cancer patients had hormone refractory or metastatic prostate cancer. Eligible lung cancer patients included those in stage IIIA, IIIB, or IV disease. Caregivers were at least 18 years of age and identified by the patient as their primary source of physical, emotional, or financial support.

Recruitment

Recruitment sites were 5 cancer centers in the Northeastern, Midwestern, and Southwestern United States. This analysis contains a combined sample from both clinical trials of 117 dyads in the CHESS-Only group and 118 in the CHESS+CR group (Figure 1). Details about recruitment, randomization, and procedures were previously reported [28,30].

The attrition rate in this study (33.6%, 36/107) in the CHESS-Only group and 38.1% (42/110) in the CHESS+CR group) is comparable with other clinical trials of patients with advanced cancer. A review of 18 interventional supportive and palliative oncology trials found an attrition rate of 44% at study end [32].

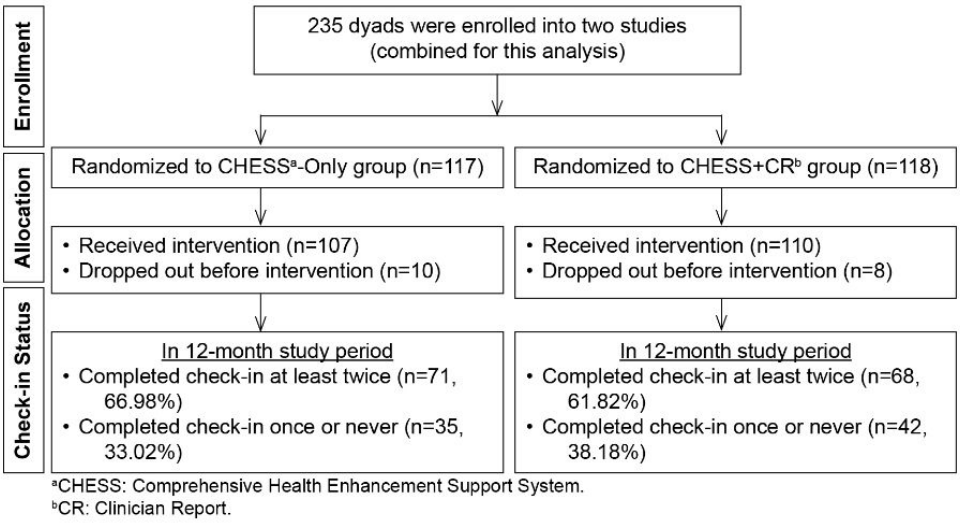
Interventions

Both the CHESS-Only and CHESS+CR participants received access to CHESS. CHESS was designed for caregivers, but patients could have access as well if they wanted it. Upon initial log-in and every 7 days after, caregivers were prompted when logging into CHESS to complete a check-in with questions about patient symptom status from a modified Edmonton Symptom Assessment Scale (ESAS) [33]. Check-in items had to be completed before advancing to another page in the CHESS website. Items on check-ins after the first check-in were populated with previous ratings; caregivers needed to respond only if a rating changed. In the CHESS-Only group, the information reported at check-in was intended for caregivers and not sent to clinicians. In the CHESS+CR group, CHESS summarized the caregiver-provided information and made it available (with patient permission) to the clinical team [31]. Alerts were emailed, faxed, or phoned (according to clinician preference) to a designated member of the clinical team, typically a nurse, (1) when a threshold symptom was reported and (2) 2 days before a scheduled clinic visit. The severity threshold for an alert was set at 7, based on Serlin et al [34], indicating that this symptom distress level interferes significantly with a patient's life. Alerts included patient name, the symptom or symptoms of concern, and worrisome symptom rating(s), along with a link to the CHESS+CR website to view the complete CR, including ratings over time. At any time, clinicians in the CHESS+CR group could also access the full CR by logging onto CHESS+CR.

Procedures

The trials were approved by the institutional review boards at each recruitment site. Oncology clinicians were consented and agreed to receive alerts.

Figure 1. Participant flow.



[a]CHESS: Comprehensive Health Enhancement Support System.
[b]CR: Clinician Report.

After caregivers and patients completed the consent form and pretest, a random number generator at the University of Wisconsin randomly assigned dyads to CHESS-Only or CHESS+CR (1:1 ratio). Randomization was blocked by dyad

relationship (spouse or partner vs nonspouse or partner) and race (white vs non-white). All caregivers who needed a computer were mailed a laptop with Internet access and a user manual. Participants who already had a computer with Internet access were reimbursed for Internet access during the study period. Caregivers (and patients, if they wished to use CHESS) were assigned a unique code name and password for accessing CHESS. Technical support was available by telephone. CHESS staff provided training on using CHESS via telephone or in clinic. Those in the CHESS+CR group were told that symptoms reported as "high" would trigger an email to the clinical team. Participants were not told the threshold, but when they gave a rating of ≥7, they were encouraged to call the clinic and notified on the website that the clinical team would be alerted.

Although caregivers and patients were enrolled as dyads, caregivers were the target population in both clinical trials. The analysis uses data from all 12 months of the breast and prostate cancer intervention and, to standardize data from the two trials, from the first 12 months of the 24-month lung cancer intervention. The CHESS website server generated a log file for storing the symptom ratings, who reported the ratings, when the ratings were reported, and whether an alert was sent. These data were retrieved and analyzed for this study.

Measures

Patient distress is a subjective measure that was assessed at check-in by a modified ESAS [30] on a 0 to 10 scale, with 10 indicating the greatest symptom distress. On the basis of the feedback from oncologists, we replaced 3 physical items in the original scale [33] (activity, drowsiness, and well-being) with three common cancer symptoms (fatigue, constipation, and diarrhea). The modified ESAS contained 10 items. This analysis focuses on individual ratings for each of the 10 symptoms rather than on a single scale score calculated across symptoms.

Improvement was determined separately for each threshold symptom by comparing subsequent ratings given by the same person at check-in. A threshold symptom was considered improved if it was rated lower at the next check-in (eg, if a caregiver rated patient pain at 9 one week and 8 the next week, the symptom was considered improved; likewise, a symptom rated at 7 one week and 6 the next was considered improved). Patient symptom ratings given in the last check-in were not examined because no follow-up check-in could be compared with it. Patients and caregivers who completed fewer than two check-ins were excluded in the analysis because they supplied no data for comparison.

To evaluate the impact of the CR on symptom change, caregiver-reported check-in data were aggregated in six 2-month periods (eg, months 3 and 4). This interval is somewhat arbitrary. It was chosen because caregivers in the two randomized trials from which the data came filled out written surveys every 2 months. The following values were calculated: (1) number of assessed symptoms (number of discrete symptoms rated during a 2-month period, calculated as the number of times a check-in was completed multiplied by 10 because each check-in assessed 10 symptoms), (2) number of patient threshold symptoms (number of symptoms rated ≥7), and (3) number of improved patient threshold symptoms (number of threshold symptoms with a lower rating in the next check-in). Group totals of the foregoing three variables were used to calculate the following two proportion indices for each 2-month period as well as the entire 12-month study period. Our primary interest was the proportion of improved caregiver-reported patient threshold symptoms (number of improved threshold symptoms out of the total number of threshold symptoms, which shows the impact on symptom management). We also calculated the proportion of caregiver-reported patient threshold symptoms (number of caregiver-reported threshold symptoms out of the total number of symptoms that were reported on). The purpose of the second proportion was to examine whether knowledge that their data might be reported to the clinical team would affect reporting behavior.

Statistical Analysis

To compare the two outcomes of interest between groups over time (ie, improved patient threshold symptoms and proportion of patient threshold symptoms), the area under the curve (AUC) was calculated based on group-aggregated values. The AUC per group was calculated using the trapezoidal rule in NCSS 2007 [35]. Group differences in AUC were assessed by converting group AUC into a relative proportion per group and then conducting a proportional difference test (StatXact 5, Cytel) [36]. In addition, aggregated symptom reports were averaged across the 12 months. Differences in these averaged improved threshold symptoms and proportions of threshold symptoms were tested using the same method. The standardized statistics, *P* value and 95% CIs, were calculated based on methods outlined by Miettinen and Nurminen [37] and Chan and Zhang [38]. To test for potential response bias after group assignment, the proportion of patient threshold symptoms reported by caregivers at pretest was compared with the proportion at the first check-in using the multiple-sample McNemar test [39]. All tests were conducted at alpha=.05 level.

Results

Baseline Demographics and ESAS Ratings

Table 1 shows demographics and pretest ESAS ratings for the two groups. Among patients, 55.8% (121/217) were female, with an average age of 63 years. Caregivers were predominantly female (64.2%, 140/217), with an average age of 56 years. Most caregivers (69.3%, 150/217) were spouses or partners. Demographic characteristics of caregivers omitted from the analysis (ie, caregivers who submitted fewer than two check-ins) were similar to those of caregivers included in the analysis except on caregiver gender, for which 53% of excluded caregivers and 71% of included caregivers were female (χ^2_1=6.4 *P*=.01). In the 12-month period, the proportion of patient symptoms reported by caregivers at least twice did not differ by randomization group (CHESS-Only, 71/107, 66.4% vs CHESS+CR, 68/110, 61.8%, standardized difference Z=0.696, *P*=.49, 95% CI −.082 to .171).

Table 1. Demographics of participants who received the interventions.

Characteristics	CHESS[a]-Only (n=107)	CHESS+CR[b] (n=110)
Cancer type, n (%)		
Breast	45 (42)	44 (40)
Prostate	30 (28)	34 (31)
Lung	32 (30)	32 (29)
Patients		
Age, mean (SD)	62.53 (9.63)[c]	62.73 (11.00)[d]
Gender, n (%)		
Male	47 (44)	49 (45)
Female	60 (56)	61 (55)
Caregivers		
Age, mean (SD)	55.73 (13.02)[c]	56.36 (13.39)[e]
Gender, n (%)		
Female	71 (66)	69 (63)
Relationship to patient, n (%)		
Spouse	75 (70)	75 (68)
Nonspouse	32 (30)	35 (32)
Caregiver annual household income in US dollars, n (%)		
Below $40,000	35 (33)	35 (32)[c]
$40,001-$80,000	37 (35)	36 (33)
$80,001 and over	26 (24)	28 (25)
Didn't report	9 (8)	11 (10)
Caregiver education (1-6)[f], mean (SD)	3.96 (1.58)[g]	3.67 (1.52)[d]
Caregiver Internet comfort (1-4)[h], mean (SD)	2.57 (1.26)[i]	2.36 (1.37)[e]
Caregiver-reported patient ESAS[j] (1-90)[k], mean (SD)	27.75 (16.82)[l]	28.13 (15.90)[i]

[a]CHESS: Comprehensive Health Enhancement Support System.

[b]CR: Clinician Report.

[c]n=105.

[d]These values were based on pretests of 109 caregivers in CHESS+CR group because 1 caregiver did not return the pretest survey.

[e]n=107.

[f]Caregiver education levels: 1 Stopped school before finishing high school; 2 High school degree; 3 Some college courses; 4 Associate or technical degree (2-year college); 5 Bachelor's degree (4 year college); 6 Graduate degree.

[g]n=106.

[h]Internet comfort levels: 0 Not at all; 1 A little; 2 A medium amount; 3 Quite a bit; 4 Extremely.

[i]n=102.

[j]ESAS: Edmonton Symptom Assessment Scale.

[k]Calculated as the sum of severity ratings (0=none; 10=worst possible) across 9 items.

[l]n=94.

[m]n=101.

Table 2. Bimonthly summary of caregiver-reported patient symptom indices. At pretest, ESAS items were rated via paper survey. The report of a threshold symptom did not produce a clinician alert.

Randomization group	CHESS[a]-Only	CHESS+CR[b]
Pretest	117	118
Month 2	101	104
Month 4	96	93
Month 6	88	85
Month 8	78	74
Month 10	72	68
Month 12	65	63

[a]CHESS: Comprehensive Health Enhancement Support System.

[b]CR: Clinician Report.

We also examined data about participant use of the system to see whether different participants used the system differently (ie, by gender, race, age, living situation, education, employment status, or income) and found no statistically significant difference.

Symptom Indices and Proportion of Threshold to All Assessed Symptoms

Table 2 shows patient symptom indices reported by caregivers calculated for each 2-month period and totaled across 12 months. Table 3 reports on the number of patient threshold symptoms reported by caregivers as a proportion of all assessed symptoms. At pretest, for instance, CHESS-Only caregivers reported on symptoms 1026 times, and 182 of those symptoms (17.74%) met or exceeded the threshold. CHESS+CR caregivers reported 1056 symptoms, and 184 (17.42%) were threshold symptoms. The effect size of the difference between CHESS-Only and CHESS+CR was 0.01. Effect sizes were determined by using Cohen arcsine transformation of the probabilities [40]. At pretest, there were no significant differences between groups, but differences emerged (caregivers in CHESS+CR reported fewer symptoms) with what Cohen describes as a small effect size.

Proportion of Improved Threshold Symptoms to All Threshold Symptoms

Table 4 reports on the number of improved threshold symptoms as a proportion of all threshold symptoms reported. For instance, at 2 months, the CHESS-Only caregivers reported 385 threshold symptoms, 103 of which (26.8%) were subsequently reported as improved. By contrast, at the same time point, CHESS+CR caregivers reported 212 threshold symptoms, 113 of which (53.3%) were subsequently reported as improved. Comparing CHESS vs CHESS+CR across all posttest time periods yielded a moderate effect size of 0.60 in favor of CHESS+CR.

Analyses of average aggregated proportions across all 12 months (the Total column of Tables 3 and 4) show similar findings: the CHESS+CR group was more likely to report improvement (53.04% vs 26.16%, Wald Z-test=10.35, $P<.001$, 95% CI .216 to .320) but less likely to report threshold symptoms (7.7% vs 14.4%, Wald Z-test=−12.27, $P<.001$, 95% CI −077 to −057). That is, throughout the study period, caregivers in the CHESS+CR group consistently reported that their patients had less symptom burden and better symptom management than patients in CHESS-Only group.

Table 3. Proportion of threshold symptoms/all assessed symptoms and effect sizes.

Randomization group	CHESS[a]-Only, n/N (%)	CHESS+CR[b], n/N (%)	Effect size (95% CI)
Pretest	182/1026 (17.74)	184/1056 (17.42)	0.01 (0.00-0.02)
Month 2	385/2620 (14.69)	212/2380 (8.91)	0.18 (0.17-0.19)
Month 4	263/1820 (14.45)	83/1240 (6.70)	0.26 (0.23-0.28)
Month 6	219/1490 (14.70)	74/840 (8.81)	0.18 (0.16-0.21)
Month 8	133/1140 (11.67)	34/550 (6.2)	0.19 (0.14-0.25)
Month 10	140/910 (15.38)	26/390 (6.7)	0.28 (0.21-0.36)
Month 12	129/830 (15.54)	14/350 (4.0)	0.41 (0.27-0.55)
Total	1269/8810 (14.40)	443/5750 (7.70)	0.216 (0.221-0.220)

[a]CHESS: Comprehensive Health Enhancement Support System.

[b]CR: Clinician Report.

Table 4. Proportion of improved threshold symptoms/all threshold symptoms and effect sizes.

Randomization group	CHESS[a]-Only, n/N (%)	CHESS+CR[c], n/N (%)	Effect size (95% CI)
Pretest	N/A[b]	N/A	N/A
Month 2	103/385 (26.8)	113/212 (53.3)	0.55 (0.52-0.57)
Month 4	67/263 (25.5)	44/83 (53)	0.57 (0.52-0.62)
Month 6	53/219 (24.2)	41/74 (55)	0.60 (0.55-0.66)
Month 8	39/133 (29.3)	17/34 (50)	0.42 (0.30-0.56)
Month 10	33/140 (23.6)	14/26 (54)	0.64 (0.48-0.79)
Month 12	37/129 (28.7)	6/14 (43)	0.30 (0.04-0.64)
Total	332/1269 (26.16)	235/443 (53.0)	0.60 (0.59-0.62)

[a]CHESS: Comprehensive Health Enhancement Support System.

[b]N/A: not applicable.

[c]CR: Clinician Report.

Examination of Possible Response Bias

Caregivers in the CHESS+CR group reported a lower proportion of threshold symptoms at each 2-month period (Table 3). It may be that they felt they might be bothering the doctor or that word would get back to the patient that the caregiver thought the patient was doing poorly. To examine whether caregivers ranked symptoms lower to avoid triggering the alert to the clinician, we compared symptom ratings at pretest with those at the first online check-in, assuming any reluctance to alert the clinician would not have been present on pretest because pretest data were not sent to clinicians and patients knew this. A multiple-sample McNemar test [39] was used to test for differences between the CHESS-Only and CHESS+CR caregiver groups on changes in reporting threshold symptoms from pretest to first online check-in. A total of 1329 symptoms were rated by 135 caregivers (70 CHESS-Only, 65 CHESS+CR) at pretest and first online check-in and were used in this analysis. At pretest, no significant difference was found between randomized groups in the number of patient threshold symptoms reported (Z=0.189, *P*=.85, 95% CI −0.03 to 0.04), but the difference in reporting threshold symptoms from pretest to first online check-in was statistically significant (Z=6.910, *P*<.001, 95% CI 6.50 to 7.29), with CHESS+CR caregivers reporting a lower proportion of threshold symptoms at first check-in than at pretest (Table 5), suggesting that CHESS+CR caregivers may have had a response bias toward lower ratings when they knew a clinician might be alerted.

Further analyses also suggest that bias may help explain the results. First, we examined the relationship between caregiver demographics and the likelihood of reporting symptom data in the CHESS-Only versus CHESS+CR groups. Gender was the only statistically significant characteristic related to the reporting of symptoms: 64.7% (172/266) of women are reporters overall versus 49.4% (131/266) of men; χ^2_1 (N=266)=5.2, *P*=.02. In the CHESS+CR group, 64.4% (75/117) of women versus 45.5% (53/117) of men were reporters, χ^2_1 (N=117)=4.02, *P*=.045; no significant differences were found in the CHESS-Only group. We also looked at 35 outcomes, such as caregiving burden and patient quality of life at each survey time frame, and did not find a consistent difference between caregivers who reported symptoms and those who did not.

We also examined, in addition to improved symptoms, which were the focus of the study, the proportion of symptoms that stayed at the same level of severity and the proportion that worsened. Caregivers in the CHESS-Only group reported more total symptoms (8810) and a greater proportion of threshold symptoms (14.40%, 1269/8810) than caregivers in the CHESS+CR group (5750 total symptoms, of which 7.70% [443/5750] were threshold symptoms). Caregivers in the CHESS+CR group reported a much higher percentage of improved threshold symptoms (53.0% [235/443] versus 26.16% [332/1269] in the CHESS-Only group), a slightly larger percentage of worsened symptoms (14.0% [62/443] versus 10.40% [132/1269] in the CHESS-Only group), and a much lower percentage of threshold symptoms with no change (33.0% [146/443] versus 63.4% [805/1269] in the CHESS-Only group).

Finally, we looked at how many caregivers reported symptoms in the 2 groups. Over the 12 months, 71 caregivers reported the 8810 symptoms in the CHESS-Only group compared with 68 caregivers in the CHESS+CR group, who reported on 5750 symptoms. On average, each reporting caregiver in the CHESS-Only group reported on 124.1 symptoms whereas caregivers in CHESS+CR group reported an average of 84.6 symptoms. Hence, a 31.8% difference exists between the groups in the average number of symptoms reported per caregiver, with those in the CHESS+CR group reporting fewer.

Table 5. Proportion of patient threshold symptoms reported by caregivers at pretest versus initial check-in.

Randomization group	CHESS[a]-Only, n/N (%)	CHESS+CR[b], n/N (%)	Effect size (95% CI)	*P* value
Pretest	106/684 (15.5)	105/645 (16.3)	0.02 (0.00-0.04)	.85
Initial check-in	108/684 (15.8)	70/645 (10.9)	0.15 (0.12-0.18)	<.001

[a]CHESS: Comprehensive Health Enhancement Support System.

[b]CR: Clinician Report.

Discussion

Principal Findings

The CR was designed to speed information about patient symptoms to clinicians by automatically sending an alert when a caregiver reported a symptom to be at or over a threshold. With immediate symptom reporting, clinicians can intervene rapidly and reduce patient symptom distress. Our results show that for symptoms causing severe (≥7 on a 0 to 10 scale) distress, patients whose caregivers had access to CHESS+CR, and therefore had an alert sent to their clinicians, had a greater proportion of symptom improvements than those with CHESS-Only, whose clinicians did not receive alerts or have access to ratings. These results suggest that the CR may facilitate patient symptom improvement and management. A previously published paper [30] found that the same system, CHESS+CR, improved caregivers' negative mood, suggesting that one electronic health (eHealth) system may help both caregivers and indirectly the patients themselves.

Possible Explanations of the Results

The effects of the CR may be explained in various ways, as clinicians explained in qualitative interviews [31], which include the following:

1. The CR could help clinicians better prepare to address patient symptoms and caregiver concerns in clinic visits.
2. The CR may boost caregiver efficacy in discussing symptoms with clinicians.
3. The CR may deepen caregiver involvement because caregivers can monitor patient symptoms and report their concerns directly to the clinical team.
4. The CR may enable earlier intervention because the severe symptom distress that triggers alerts might otherwise be unreported and therefore not attended to.

The overall assessed symptom rate in the CHESS+CR group dropped significantly more than in the CHESS-Only group. A response bias may have occurred if caregivers avoided using the check-in or rated symptoms lower because they feared bothering the clinician [41] or upsetting the patient. The finding that symptom distress was equal between the two study groups at pretest but lower in the CHESS+CR group at first check-in supports this explanation. The examination of other factors that may account for the difference between the groups in caregiver-reported symptoms (ie, demographic differences between caregivers in the two groups, the proportion of symptoms that stayed the same and worsened in each group, and the number of caregivers reporting symptoms in each group) corroborates the suggestion of bias among the CHESS+CR caregivers to report fewer symptoms.

The lower overall assessed symptom rate in the CHESS+CR group may also be explained by clinician response. Clinicians were given no directions about how to respond to CR alerts. If clinicians did not respond (we do not know whether they did), participants may have stopped using the CR, although the data suggest that when the CR alerts the clinician, the symptoms are more likely to improve. It may be that clinicians responded to alerts by addressing the symptoms more promptly, scheduling additional patient visits, or responding differently to patients during visits, though we lack the data to explore this. Future research on caregiver and patient motives for using or not using a symptom reporting system could better inform system development and dissemination.

Limitations

The study has limitations. Data collection was completed in 2009. We do not see this as a major limitation because the influence of caregiver input on clinical decision making is rarely examined, and the result of such input, as shown in this paper, can be significant for patients. In fact, because reports based on the regular collection of both patient and caregiver data are beginning to appear more frequently, the methods and results reported here are more likely to be timely now than several years ago.

Each institution had a small number of clinicians, and they were not randomized. The effect of the CR on symptom improvement could potentially be greater than demonstrated here because clinicians seeing patients in both groups may have improved their symptom management with control-group patients as a result of changes they made with the intervention group. Future research could avoid this by randomizing clinics or clinicians rather than caregiver/patient dyads within clinics, although this approach could confound clinic variability in care with randomization.

Furthermore, all outcome variables were self-reported. Medical records could have been used to validate self-reports, for example, by observing whether interventions seemed to respond to reported threshold symptoms, such as a change in the type or amount of medication if high pain levels were reported. Such an analysis was outside the scope of the study.

Although clinicians participated in qualitative interviews [31] (results are summarized above under Possible Explanation of the Results), a distinct weakness of the paper is the lack of qualitative evidence from caregivers and patients. For example, qualitative assessment could also have been used to explore the apparent bias in the CHESS+CR group about "bothering" the doctor.

Although the study suggests the potential of CR-like systems to enhance patient care and speed recovery from distressing

symptoms, further research with different patient populations would help validate and improve the generalizability of these findings. In addition, widespread use of such a system poses important challenges—cost, risk aversion, clinician time, and interoperability with the electronic health record (EHR). Caregivers could use such a system on their smartphones, which are becoming ubiquitous. Yet maintaining such an eHealth system has fixed costs (eg, for updating content and moderating discussions). These costs could be borne by health systems and insurers if outcomes warrant it, given the new financing models that reward governmental agencies to pay for systems that improve outcomes and the increasing use of fixed payments to providers. The human tendency to avoid risk and stay with the familiar works against innovations such as CR-like systems, as does the time pressure clinicians contend with. Because clinicians worry about innovations that add work and reduce time with patients, systems must be built to minimize burden. Finally, as we have discovered in subsequent studies, getting information from such a system into the EHR can be extremely difficult. However, this is important and should be done so that clinicians can access information from patient and caregiver eHealth systems without going to a website outside the EHR.

Conclusions

The significance of this study is its finding that eHealth support helped caregivers play a more effective role in their loved one's care than the role they played without that support. Furthermore, the results suggest that eHealth alerts coming from family caregivers can influence clinician behavior. Together, these findings suggest that eHealth-based CRs from caregivers can influence care for patients with many types of cancer, as well as chronic conditions such as metabolic syndrome, addiction, human immunodeficiency virus, and Alzheimer disease [42,43]. Further research is required to address this speculation.

Acknowledgments

This research was funded by two National Institutes of Health (NIH) grants (from the National Cancer Institute [1 P50 CA095817-01A1] and the National Institute of Nursing Research [RO1 NR008260-01]) and one Agency for Healthcare Research and Quality grant (5P50HS019917-04). The funders had no role in any aspect of the development, conduct, analysis, or reporting of the study.

Conflicts of Interest

DHG, FM, RAJ, and AQ have a shareholder interest in CHESS Mobile Health, a public benefit corporation that develops health care technology for patients struggling with addiction. This relationship is managed by the authors and the University of Wisconsin-Madison's Conflict of Interest Committee.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V.1.6.1).

[PDF File (Adobe PDF File), 675KB - jmir_v19i11e354_app1.pdf]

References

1. Glajchen M. The emerging role and needs of family caregivers in cancer care. J Support Oncol 2004;2(2):145-155. [Medline: 15328817]
2. Haley WE. Family caregivers of elderly patients with cancer: understanding and minimizing the burden of care. J Support Oncol 2003;1(4 Suppl 2):25-29. [Medline: 15346997]
3. Given BA, Given CW, Kozachik S. Family support in advanced cancer. CA Cancer J Clin 2001;51(4):213-231 [FREE Full text] [Medline: 11577488]
4. Walling AM, Keating NL, Kahn KL, Dy S, Mack JW, Malin J, et al. Lower patient ratings of physician communication are associated with unmet need for symptom management in patients with lung and colorectal cancer. J Oncol Pract 2016 Jun;12(6):e654-e669. [doi: 10.1200/JOP.2015.005538] [Medline: 27221991]
5. Twycross R, Wilcock A. Symptom management in advanced cancer. 3rd ed. London: Radcliffe Medical Press; 2002.
6. Bloche MG. Managing conflict at the end of life. N Engl J Med 2005 Jun 09;352(23):2371-2373. [doi: 10.1056/NEJMp058104] [Medline: 15944420]
7. Van Keer R, Deschepper R, Francke AL, Huyghens L, Bilsen J. Conflicts between healthcare professionals and families of a multi-ethnic patient population during critical care: an ethnographic study. Crit Care 2015 Dec 22;19:441 [FREE Full text] [doi: 10.1186/s13054-015-1158-4] [Medline: 26694072]
8. Gustafson DH. A good death. J Med Internet Res 2007 Mar 14;9(1):e6 [FREE Full text] [doi: 10.2196/jmir.9.1.e6] [Medline: 17478415]
9. Parker-Pope T. Well.blogs.nytimes. New York: The New York Times; 2008. 'Bothering' Your Doctor URL: https://well.blogs.nytimes.com/2008/03/17/bothering-your-doctor/ [accessed 2017-06-29] [WebCite Cache ID 6o0bdUbYX]
10. Prentice JC, Pizer SD. Delayed access to health care and mortality. Health Serv Res 2007 Apr;42(2):644-662 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00626.x] [Medline: 17362211]

11. Arno PS, Levine C, Memmott MM. The economic value of informal caregiving. Health Aff (Millwood) 1999;18(2):182-188 [FREE Full text] [Medline: 10091447]
12. Apple, Inc. Apple. Cupertino, CA: Apple, Inc; 2017. A bold new way to look at your health URL: http://www.apple.com/ios/health/ [accessed 2017-06-29] [WebCite Cache ID 6nvpMO7Rc]
13. Symple Health, Inc. itunes.apple. Cupertino, CA: iTunes, Apple, Inc; 2017. Symple - Symptom Tracker & Health Diary URL: https://itunes.apple.com/us/app/symple-symptom-tracker-health/id479818115?mt=8 [accessed 2017-06-29] [WebCite Cache ID 6nvqLZmQR]
14. Coons SJ, Eremenco S, Lundy JJ, O'Donohoe P, O'Gorman H, Malizia W. Erratum to: Capturing Patient-Reported Outcome (PRO) data electronically: the past, present, and promise of ePRO measurement in clinical trials. Patient 2015 Dec;8(6):571 [FREE Full text] [doi: 10.1007/s40271-015-0121-4] [Medline: 25701173]
15. Cleeland CS, Wang XS, Shi Q, Mendoza TR, Wright SL, Berry MD, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. J Clin Oncol 2011 Mar 10;29(8):994-1000 [FREE Full text] [doi: 10.1200/JCO.2010.29.8315] [Medline: 21282546]
16. Ruland CM, Holte HH, Røislien J, Heaven C, Hamilton GA, Kristiansen J, et al. Effects of a computer-supported interactive tailored patient assessment tool on patient care, symptom distress, and patients' need for symptom management support: a randomized clinical trial. J Am Med Inform Assoc 2010;17(4):403-410 [FREE Full text] [doi: 10.1136/jamia.2010.005660] [Medline: 20595307]
17. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
18. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724 [FREE Full text] [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
19. Gilbert JE, Howell D, King S, Sawka C, Hughes E, Angus H, et al. Quality improvement in cancer symptom assessment and control: the Provincial Palliative Care Integration Project (PPCIP). J Pain Symptom Manage 2012 Apr;43(4):663-678. [doi: 10.1016/j.jpainsymman.2011.04.028] [Medline: 22464352]
20. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. J Am Med Assoc 2017 Jun 04;18(2):197-198. [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
21. Nekolaichuk CL, Bruera E, Spachynski K, MacEachern T, Hanson J, Maguire TO. A comparison of patient and proxy symptom assessments in advanced cancer patients. Palliat Med 1999 Jul;13(4):311-323. [Medline: 10659100]
22. Lobchuk MM, Degner LF. Symptom experiences: perceptual accuracy between advanced-stage cancer patients and family caregivers in the home care setting. J Clin Oncol 2002 Aug 15;20(16):3495-3507. [doi: 10.1200/JCO.2002.01.153] [Medline: 12177111]
23. Kutner JS, Bryant LL, Beaty BL, Fairclough DL. Symptom distress and quality-of-life assessment at the end of life: the role of proxy response. J Pain Symptom Manage 2006 Oct;32(4):300-310. [doi: 10.1016/j.jpainsymman.2006.05.009] [Medline: 17000347]
24. Gustafson D, Wise M, Bhattacharya A, Pulvermacher A, Shanovich K, Phillips B, et al. The effects of combining Web-based eHealth with telephone nurse case management for pediatric asthma control: a randomized controlled trial. J Med Internet Res 2012;14(4):e101 [FREE Full text] [doi: 10.2196/jmir.1964] [Medline: 22835804]
25. Gustafson DH, Hawkins R, McTavish F, Pingree S, Chen WC, Volrathongchai K, et al. Internet-based interactive support for cancer patients: are integrated systems better? J Commun 2008 Jun;58(2):238-257 [FREE Full text] [doi: 10.1111/j.1460-2466.2008.00383.x] [Medline: 21804645]
26. Gustafson DH, Hawkins RP, Boberg EW, McTavish F, Owens B, Wise M, et al. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. Int J Med Inform 2002 Nov 12;65(3):169-177. [Medline: 12414016]
27. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
28. Gustafson DH, DuBenske LL, Namkoong K, Hawkins R, Chih M, Atwood AK, et al. An eHealth system supporting palliative care for patients with non-small cell lung cancer: a randomized trial. Cancer 2013 May 1;119(9):1744-1751 [FREE Full text] [doi: 10.1002/cncr.27939] [Medline: 23355273]
29. DuBenske LL, Gustafson DH, Shaw BR, Cleary JF. Web-based cancer communication and decision making systems: connecting patients, caregivers, and clinicians for improved health outcomes. Med Decis Making 2010;30(6):732-744 [FREE Full text] [doi: 10.1177/0272989X10386382] [Medline: 21041539]
30. Chih M, DuBenske LL, Hawkins RP, Brown RL, Dinauer SK, Cleary JF, et al. Communicating advanced cancer patients' symptoms via the Internet: a pooled analysis of two randomized trials examining caregiver preparedness, physical burden, and negative mood. Palliat Med 2013 Jun;27(6):533-543 [FREE Full text] [doi: 10.1177/0269216312457213] [Medline: 22988042]

31. Dubenske LL, Chih M, Dinauer S, Gustafson DH, Cleary JF. Development and implementation of a clinician reporting system for advanced stage cancer: initial lessons learned. J Am Med Inform Assoc 2008;15(5):679-686 [FREE Full text] [doi: 10.1197/jamia.M2532] [Medline: 18579837]
32. Hui D, Glitza I, Chisholm G, Yennu S, Bruera E. Attrition rates, reasons, and predictive factors in supportive care and palliative oncology clinical trials. Cancer 2013 Mar 01;119(5):1098-1105 [FREE Full text] [doi: 10.1002/cncr.27854] [Medline: 23132290]
33. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. J Palliat Care 1991;7(2):6-9. [Medline: 1714502]
34. Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. Pain 1995 May;61(2):277-284. [Medline: 7659438]
35. Hintze J. NCSS-wpengine. Kaysville, UT: Number Cruncher Statistical Systems Software (NCSS), LLC; 2007. NCSS 2007 user guide URL: https://ncss-wpengine.netdna-ssl.com/wp-content/uploads/2012/09/NCSSUG1.pdf [accessed 2017-08-09] [WebCite Cache ID 6sb24m4jD]
36. Mehta C, Patel N. StatXact 5. Cambridge, MA: Cytel Software Corporation; 2001.
37. Miettinen O, Nurminen M. Comparative analysis of two rates. Stat Med 1985;4(2):213-226. [Medline: 4023479]
38. Chan IS, Zhang Z. Test-based exact confidence intervals for the difference of two binomial proportions. Biometrics 1999 Dec;55(4):1202-1209. [Medline: 11315068]
39. Marascuilo L, Serlin R. Tests and contrasts for comparing change parameters for a multiple sample McNemar data model. Br J Math Stat Psychol 1979;32(1):105-112. [doi: 10.1111/j.2044-8317.1979.tb00755.x]
40. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
41. Worth A, Boyd K, Kendall M, Heaney D, Macleod U, Cormie P, et al. Out-of-hours palliative care: a qualitative study of cancer patients, carers and professionals. Br J Gen Pract 2006 Jan;56(522):6-13 [FREE Full text] [Medline: 16438809]
42. Haley WE, Brown SL, Levine EG. Family caregiver appraisals of patient behavioral disturbance in senile dementia. Clin Gerontol 1987 Sep 30;6(4):25-34. [doi: 10.1300/J018v06n04_04]
43. Schulz R, Martire LM. Family caregiving of persons with dementia: prevalence, health effects, and support strategies. Am J Geriatr Psychiatry 2004;12(3):240-249 [FREE Full text] [Medline: 15126224]

Abbreviations

AUC: area under the curve
CHESS: Comprehensive Health Enhancement Support System
CR: Clinician Report
eHealth: electronic health
ESAS: Edmonton Symptom Assessment System
EHR: electronic health record
NIH: National Institutes of Health

Edited by G Eysenbach; submitted 08.02.17; peer-reviewed by T Ginossar, N Goonawardene; comments to author 28.03.17; revised version received 10.05.17; accepted 27.07.17; published 14.11.17.

Please cite as:
Gustafson DH, DuBenske LL, Atwood AK, Chih MY, Johnson RA, McTavish F, Quanbeck A, Brown RL, Cleary JF, Shah D
Reducing Symptom Distress in Patients With Advanced Cancer Using an e-Alert System for Caregivers: Pooled Analysis of Two Randomized Clinical Trials
J Med Internet Res 2017;19(11):e354
URL: http://www.jmir.org/2017/11/e354/
doi:10.2196/jmir.7466
PMID:29138131

Original Paper

A Fully Automated Web-Based Program Improves Lifestyle Habits and HbA1c in Patients With Type 2 Diabetes and Abdominal Obesity: Randomized Trial of Patient E-Coaching Nutritional Support (The ANODE Study)

Boris Hansel[1,2,3], MD; Philippe Giral[4,5,6], MD, PhD; Laetitia Gambotti[7], MD, PhD; Alexandre Lafourcade[7], PhD; Gilbert Peres[6], MD, PhD; Claude Filipecki[8], MD; Diana Kadouch[1], MD; Agnes Hartemann[5,6,9], PhD, MD; Jean-Michel Oppert[5,6,8], MD, PhD; Eric Bruckert[4,5,6], MD; Michel Marre[1,2,3], MD, PhD; Arnaud Bruneel[10], MD; Emilie Duchene[5], PEng; Ronan Roussel[1,2,3,11], MD, PhD

[1]Department of Endocrinology, Diabetology, Nutrition, Hôpitaux Universitaires Paris-Nord Val de Seine, Assistance Publique-Hôpitaux de Paris, Paris, France

[2]Paris Diderot Université, Sorbonne Paris Cité, Paris, France

[3]Centre de Recherche des Cordeliers, Institut National de la Santé et de la Recherche Médicale, Paris, France

[4]Department of Endocrinology, Hôpital Pitié-Salpêtrière, Paris, France

[5]Institute of Cardiometabolism and Nutrition, Paris, France

[6]Université Pierre et Marie Curie, Université de Paris, Sorbonne Universités, Paris, France

[7]L'Unité de Recherche Clinique, Hôpital Pitié-Salpêtrière, Paris, France

[8]Department of Nutrition, Hôpital Pitié-Salpêtrière, Paris, France

[9]Department of Diabetology, Hôpital Pitié-Salpêtrière, Paris, France

[10]Department of Biochemistry, Hôpitaux Universitaires Paris-Nord Val de Seine, Assistance Publique-Hôpitaux de Paris, Paris, France

[11]Département Hospitalo-Universitaire, Fibrosis, Inflammation, Remodelling, Paris, France

Corresponding Author:
Boris Hansel, MD
Hôpitaux Universitaires Paris-Nord Val de Seine
Department of Endocrinology, Diabetology, Nutrition
Assistance Publique-Hôpitaux de Paris
46 Rue Henri Huchard
Paris, 75018
France
Phone: 33 0611307469
Email: boris.hansel@aphp.fr

Abstract

Background: The prevalence of abdominal obesity and type 2 diabetes mellitus (T2DM) is a public health challenge. New solutions need to be developed to help patients implement lifestyle changes.

Objective: The objective of the study was to evaluate a fully automated Web-based intervention designed to help users improve their dietary habits and increase their physical activity.

Methods: The Accompagnement Nutritionnel de l'Obésité et du Diabète par E-coaching (ANODE) study was a 16-week, 1:1 parallel-arm, open-label randomized clinical trial. Patients with T2DM and abdominal obesity (n=120, aged 18-75 years) were recruited. Patients in the intervention arm (n=60) had access to a fully automated program (ANODE) to improve their lifestyle. Patients were asked to log on at least once per week. Human contact was limited to hotline support in cases of technical issues. The dietetic tool provided personalized menus and a shopping list for the day or the week. Stepwise physical activity was prescribed. The control arm (n=60) received general nutritional advice. The primary outcome was the change of the dietary score (International Diet Quality Index; DQI-I) between baseline and the end of the study. Secondary endpoints included changes in body weight, waist circumference, hemoglobin A1c (HbA1c) and measured maximum oxygen consumption (VO2 max).

Results: The mean age of the participants was 57 years (standard deviation [SD] 9), mean body mass index was 33 kg/m² (SD 4), mean HbA1c was 7.2% (SD 1.1), and 66.7% (80/120) of participants were women. Using an intention-to-treat analysis, the DQI-I score (54.0, SD 5.7 in the ANODE arm; 52.8, SD 6.2 in the control arm; *P*=.28) increased significantly in the ANODE arm compared to the control arm (+4.55, SD 5.91 vs -1.68, SD 5.18; between arms *P*<.001). Body weight, waist circumference, and HbA1c changes improved significantly in the intervention.

Conclusions: Among patients with T2DM and abdominal obesity, the use of a fully automated Web-based program resulted in a significant improvement in dietary habits and favorable clinical and laboratory changes. The sustainability of these effects remains to be determined.

Trial Registration: ClinicalTrials.gov NCT02343107; http://clinicaltrials.gov/ct2/show/NCT02343107 (Archived by WebCite at http://www.webcitation.org/6uVMKPRzs)

(J Med Internet Res 2017;19(11):e360) doi:10.2196/jmir.7947

KEYWORDS

e-health; nutrition; type 2 diabetes

Introduction

Obesity and type 2 diabetes mellitus (T2DM) are public health issues that are growing worldwide, and their incidence rates are closely correlated [1]. Abdominal obesity is a risk factor for T2DM, but is also an independent risk factor for cardiovascular disease; the hypothetical underlying mechanisms are an abnormal adipokine profile and increased insulin resistance [2]. Lifestyle changes that are designed to achieve a healthy diet, weight reduction, and increased physical activity are the cornerstone of the treatment of obese patients with T2DM [3]. Experts recommend individually tailored care for these patients [3]. However, a high-intensity, multidisciplinary intervention (as recommended) is often impossible to implement in real life environments due to limited human resources and the high costs of long-term care. In addition, geographically isolated patients cannot easily access face-to-face education programs. Therefore, it is necessary to develop innovative approaches to improve the adoption of a healthy lifestyle.

Remote counseling using innovative technologies is a promising tool to provide advice and monitor progress at a lower cost than face-to-face education when extended follow-up is needed. A recent review identified 13 randomized trials assessing the value of remote e-coaching for patients with T2DM [4]. The results were generally positive with programs that included an intervention on physical activity, nutrition (individualized goals based on dietary recommendations for people with T2DM and obesity), self-monitoring, or weight loss. However, most of the programs assessed included important human support. In addition, the studies did not comprehensively measure the effects on dietary habits, physical activity, and metabolic parameters. Fully automated Web-based interventions could more easily be proposed on a large scale compared to telehealth programs that require expensive human support. However, few randomized studies (consisting mainly of sending short messages) have comprehensively examined the effects of such programs for T2DM [5,6].

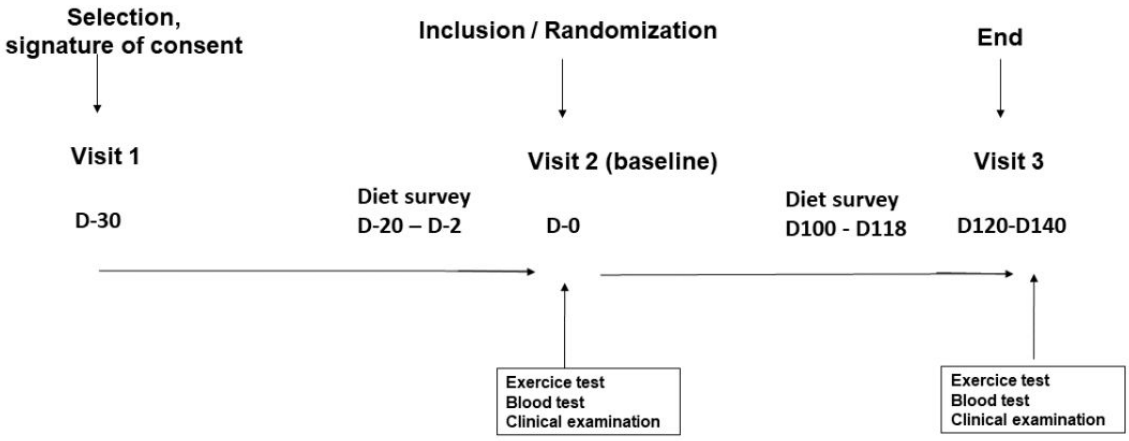
The Accompagnement Nutritionnel de l'Obésité et du Diabète par E-coaching (ANODE) tool is a fully automated Web-based program designed to help users improve their dietary habits and increase their physical activity, with an expected moderate 3-5% weight loss. ANODE includes an interactive two-way program that supports compliance with guidelines for management of obesity and T2DM. In this study, we assessed the efficacy of the ANODE program in patients with T2DM and abdominal obesity by assessing not only changes in dietary habits and physical activity, but also changes in cardiometabolic risk factors. Consequently, the primary endpoint was a quality dietary score, and secondary endpoints were parameters related to metabolism and aerobic fitness.

Methods

Study Design and Patients

The ANODE study was a randomized, comparative effectiveness clinical trial conducted between March 2014 and December 2015 in two university hospitals in Paris, France. Participants were recruited via media advertising or referred directly by their caregivers. Male or female subjects, aged 18-75 years, with abdominal obesity (defined as waist circumference >102 cm for men and >88 cm for women) and T2DM with hemoglobin A1c (HbA1c) >5.6% and <8.5% at the screening visit were included. In addition, patients were eligible if they had already received standard nutritional education for the treatment of diabetes, and if they had been on a stable, nonrestrictive diet for the past 3 months (weight change <4 kg peak-to-trough). Antidiabetic, antihypertensive, and lipid-lowering therapies had to be stable for at least 3 months. Participants had to be covered by French national health insurance. Internet access with frequent use (at least three times per week) was required, as well as an email address and fluent understanding of written and spoken French language. Patients with symptomatic cardiovascular disease, patients requiring rapid control of diabetes, those receiving general or local treatment likely to interfere with assessment of the primary endpoint, and subjects with any severe or acute illness likely to influence the results of the study (or possibly be life-threatening in the short term) were excluded. Patients who had undergone obesity surgery and those with an initial calorie intake <1200 kcal/24 hours or >4000 kcal/24 hour were also excluded. Participants in both arms were invited to attend a screening visit (first visit), an inclusion plus randomization visit (second visit), and a third and final visit at 4 months (Figure 1).

Figure 1. Study design.



The second and third visits included a clinical examination, an exercise test, and fasting blood collection. Patients provided their written informed consent. The study was approved by the Ethics Committee, Comité de Protection des Personnes Ile-de-France VI, in Paris, France.

Randomization

Randomization was performed by means of a computerized randomization program accessed via a secure Web interface. Patients were randomly assigned in a 1:1 ratio to the intervention or control arms. The randomization list was computer-generated and balanced by blocks of undisclosed size. Patients were informed of their allocation by telephone and patients allocated to the intervention arm were given a personal code to connect to the e-coaching program. Control subjects were asked to continue their usual follow-up with their general practitioner and/or specialist.

Intervention

The ANODE e-coaching program is a Web-based nutritional support tool developed and provided by the MXS Company. ANODE is designed to improve lifestyle habits, including both diet and physical activity, and consists of four modules: (1) diet and physical activity self-monitoring module, (2) nutritional assessment, (3) balanced diet menu generator, and (4) physical activity education and prescription program. Participants had to complete the questionnaires in the diet and physical activity self-monitoring module for one week to access the other three modules. Participants then accessed the four modules simultaneously, and were asked to connect at least once per week. Human contact was limited to hotline support in cases of technical issues. Patients in the intervention group had unlimited access to the ANODE program, free of charge. No monetary incentives were offered. Outlines of the ANODE e-coaching program are presented in Multimedia Appendix 1**.** At the time of the study, the program was designed to run only on personal computers.

Diet and Physical Activity Self-Monitoring Module

The aim of this module was first to collect information on participants' habits but also to keep them aware of their dietary intake and of their physical activity level. The computerized dietary survey has been used in epidemiological studies such as the French National Nutrition and Health Study [7], and was designed for self-administration based on a secured user-friendly interface, as described by Lassale et al [8]. Participants reported all foods and beverages (type and quantity) consumed during 24 hours from midnight to midnight. Participants first entered a list of every food item consumed that they could recall via one of the following two options: a food browser (foods grouped by category) or a search engine that accepts spelling errors. Participants then estimated portion sizes of foods with the help of pictures, derived from a previously validated booklet which represents more than 250 generic foods (corresponding to more than 2000 specific food items), each proposed in three different portion sizes. With two intermediate and two extreme quantities, there are seven choices of amounts. Participants could also directly enter the quantity of foods consumed in grams or a measure of volume; they could also use units of purchase (eg, one piece, one bottle, one bar) or standard household units (eg, teaspoons and tablespoons). Physical activity was collected as the daily number of steps measured with a pedometer, and as the self-reported duration and frequency of endurance activities. Participants were encouraged to enter information for each day of the week in order to further personalize the program (see prescription program below).

Nutritional Assessment

Based on 24-hour dietary recall, the program informed the patients about the mean level of calories ingested as well as the mean fat, saturated fat, protein, salt, and carbohydrate contents of their diet. Intakes of certain food groups (ie, fish, starchy foods, high-fat foods, dairy products, alcoholic beverages, and water) were also reported [8]. The program also provided advice to ensure a balanced diet according to national guidelines. The more the subject completed the dietary survey (diet and physical activity self-monitoring module), the more the nutritional assessment was precise and advices were personalized.

Balanced Diet Menu Generator

This innovative program can construct daily or weekly menus complying with the recommendations of the National Nutrition and Health Program. These menus were: (1) customized to meet the user's preferences, tastes in food, caloric level, and needs; and (2) practical, as the tool proposed daily or weekly menus adapted to the season and the selected price range (3 levels). Daily or weekly shopping lists and recipes were proposed.

Participants were encouraged to use the generator at least three times per week. However, even if he/she did not precisely follow the proposed meals, he/she was advised to draw inspiration from the proposed menus.

Physical Activity Education and Prescription Program

The objectives of this program were defined according to a stepwise approach. Five different videos explained how to perform the recommended physical activity: (1) how to perform exercise at the optimal intensity, (2) how to use a pedometer, (3) description of the 5 categories of exercise recommended in the program, (4) description of the intensity of physical activity, and (5) the step test. Each video lasted approximately 2 minutes and was accompanied by a technical sheet. If data were completely collected (see *diet and physical activity self-monitoring module* above) the program was able to fix adjusted quantitative objectives for the number of steps and for the volume (number and duration of the sessions of exercise) of endurance activities. The types of exercise (eg, cycling, brisk walking, running) and the days of the sessions were determined by the participant.

Control Arm

Participants assigned to the control arm were asked to continue their usual follow-up with their general practitioner and/or specialist.

Trial Procedures and Outcomes

After the selection visit, all participants were assessed at baseline (randomization visit) and 4 months later at a final visit.

Diet and Physical Activity Assessment

Diet was evaluated by a 3-day dietary recall (two weekdays and one weekend day). This assessment was performed before the randomization visit and during the two weeks preceding the final visit. The dates of the 3-day dietary recall were determined randomly in advance by the study team dietician. When one or more of the 3-day dietary recalls could not be completed, another three dates were randomly determined. Dietary recalls were performed with an interactive Web-based self-administered dietary diary [9] but data were verified and completed by phone with a dietician blinded to the patient's allocation. The primary endpoint of the study was the change in the dietary score (International Diet Quality Index, DQI-I [10,11]) between baseline and the end of the study. The DQI-I score (range 0-100) comprises four components (variety, adequacy, moderation, and overall balance) and the cutoffs used for adequacy and moderation were those corresponding to French recommendations [12]. Table 1 summarizes the characteristics of the DQI-I score; detailed information on computation of the DQI-I dietary score can be found elsewhere [11]. The amount of physical activity was evaluated using the short version of the International Physical Activity Questionnaire (IPAQ) [13] completed online by the subjects before randomization and at the final visit. Due to a technical issue, sitting time (although part of the IPAQ questionnaire) was not recorded during the trial.

Clinical Measurements

Body weight was measured to the nearest 0.1 kg, with participants fasting and clothed without shoes, using automated digital scales (TANITA T6360, Tanita Co., Tokyo, Japan). Body mass index (BMI) was calculated as the ratio between weight and height squared with weight expressed in kilograms and height expressed in meters. Waist circumference was measured with an inelastic tape to the nearest 0.1 cm at the level of the iliac crest during shallow breathing.

Laboratory Assessments

Total cholesterol, high-density lipoprotein-cholesterol (HDL-C) and triglyceride concentrations were determined by automated enzymatic methods. Low-density lipoprotein (LDL-C) was calculated using Friedewald's equation for triglycerides <340 mg/dl (3.9 mmol/L), or measured directly for triglycerides >340 mg/dl (3.9 mmol/L) but <700 mg/dl (8 mmol/L). Blood samples were drawn for assays of creatinine, uric acid, liver enzymes, fasting glucose, and HbA1c as part of routine care. Systemic inflammation was assessed by the plasma concentration of high-sensitivity C-reactive protein (hs-CRP).

Measurement of Aerobic Fitness

Aerobic fitness was determined by maximum oxygen consumption (VO_2 max), measured by an incremental cardiopulmonary exercise test on a cycle ergometer, with increments of 15-30 Watts/2 minutes [14]. Oxygen consumption was expressed as VO_2 per kg total body weight per minute (mL/kg/minute). We tried to achieve maximal cardiopulmonary exercise testing by using strong verbal encouragements. After exhaustion, active then passive recoveries were recorded with the patient seated.

Table 1. Characteristics and computation of the DQI-I diet quality scores [10,11].

Components	Score range: 0-100 points
Variety	Overall food group variety (0-15 points); within-group variety for protein source (0-5 points)
Adequacy	Vegetables, fruits, cereals, fiber, protein, iron, calcium, vitamin C (0-5 points each); nutritional recommendations are country-specific (in this case France)
Moderation	Total fat, saturated fat, cholesterol, sodium, empty-energy foods (0-6 points each)
Overall balance	Macronutrient ratio (carbohydrate : protein : fat, 0-6 points); fatty acid ratio (polyunsaturated fatty acid : monounsaturated fatty acid : saturated fatty acid, 0-4 points)

Satisfaction Questionnaire

In a post hoc analysis, we collected some information about the subjective opinion of the patients in the intervention group. The questionnaire was sent by email. Patients were asked if they would recommend the program and if they encountered difficulties with the use of the platform. In addition, respondents were encouraged to send comments about the quality of the program.

Required Number of Subjects

To show a 3-point difference in the variation of the DQI-I scores between the two arms over a 4-month period (with an SD of 5, a statistical power of 80%, and a type 1 error of 5% using a two-tailed test), the required number of subjects was estimated to be 45 per arm. Sixty subjects were included in each arm to anticipate patients lost to follow-up.

Statistical Analysis

Baseline characteristics in each study arm were expressed as frequencies and percentages for categorical variables, and as means and standard deviations for continuous variables. Changes in DQI-I scores (primary outcome) were compared between the two randomization arms using student t-tests. In cases of missing data, changes in DQI-I score were imputed to 0 (no variation). Other continuous variables were compared using student t-tests or Wilcoxon tests, and categorical variables were compared using the Chi-square or Fisher's exact tests, as appropriate. To assess whether the effect of the ANODE program on the DQI-I score varied according to baseline characteristics, post hoc subgroup analyses were performed according to age, sex, HbA1c, BMI, education level, and occupation. In these analyses, continuous variables were dichotomized according to the median, and interaction tests were performed. Finally, in addition to the intention-to-treat analysis, a per-protocol analysis was also performed, using only complete data. All analyses were conducted with the two-sided alpha risk of 5%, and were performed using SAS software version 9.2.

Results

Participants

Figure 2 shows the flow chart of participants. Among the volunteers who contacted the study team for participating in the study, 363 were subsequently contacted by phone and 137 were considered eligible. After the exclusion of 17 subjects who eventually did not meet all criteria (n=16) or who finally decided not to participate (n=1), 120 subjects were enrolled and randomized. At the end of the study, 13 subjects (8 in the intervention arm and 5 in the control arm) were unable to be contacted to collect dietary data, and 17 subjects (12 in the intervention arm and 5 in the control arm) did not attend the last visit. The primary endpoint was available for 107 of 120 (89.2%) randomized subjects.

Baseline characteristics were similar between the two arms (Multimedia Appendix 2). The majority of participants were females (80/120, 66.7%). On average, participants were moderately obese with no difference in baseline DQI-I scores. Intake of the main nutrients was comparable between arms.

Dietary Changes

Dietary characteristics of the participants are presented in Multimedia Appendix 2. The DQI-I score was 54.0 (SD 5.7) in the e-coaching arm and 52.8 (SD 6.2) in the control arm (P=.28 between arms). The intention-to-treat analysis at 16 weeks showed that e-coaching resulted in a significant improvement in the DQI-I score: +4.55 (SD 5.91) in the intervention arm and -1.68 (SD 5.18) in the control arm (P<.01 between arms; Table 2).

Moreover, changes in dietary intake tended to differ between arms for lipids (P=.02), saturated fats (P<.01), sodium (P=.07), and empty calories (P=.06), always towards healthier foods in the intervention arm (Table 3).

Changes in Anthropometric Variables, Cardiometabolic Risk Factors, and Aerobic Fitness

Compared to usual care, the ANODE program was associated with reduced body weight, waist circumference, and HbA1c (Multimedia Appendix 2). A significantly higher proportion of e-coaching subjects achieved weight loss >3% (20/60, 33.3% of e-coaching subjects vs 4/60, 6.7% of control subjects on intention-to-treat analysis; P<.01) and weight loss >5% (12/60, 20.0% vs 2/60, 3.3% on intention-to-treat analysis; P<.01; Figure 3). Only two subjects (2/60, 3.3%) in the ANODE arm and no subject in the control arm achieved >10% weight loss, respectively (P=.15 between arms).

No significant differences in terms of change in blood pressure, plasma lipids, aminotransferases, gamma glutamyl aminotransferase, uric acid, fasting glucose, VO_2 max or hs-CRP were observed between the two arms at 4 months (Table 4). No significant changes in physical activity assessed by the IPAQ questionnaire were observed in the intervention arm compared to control subjects (data not shown).

Figure 2. Flowchart of the study.

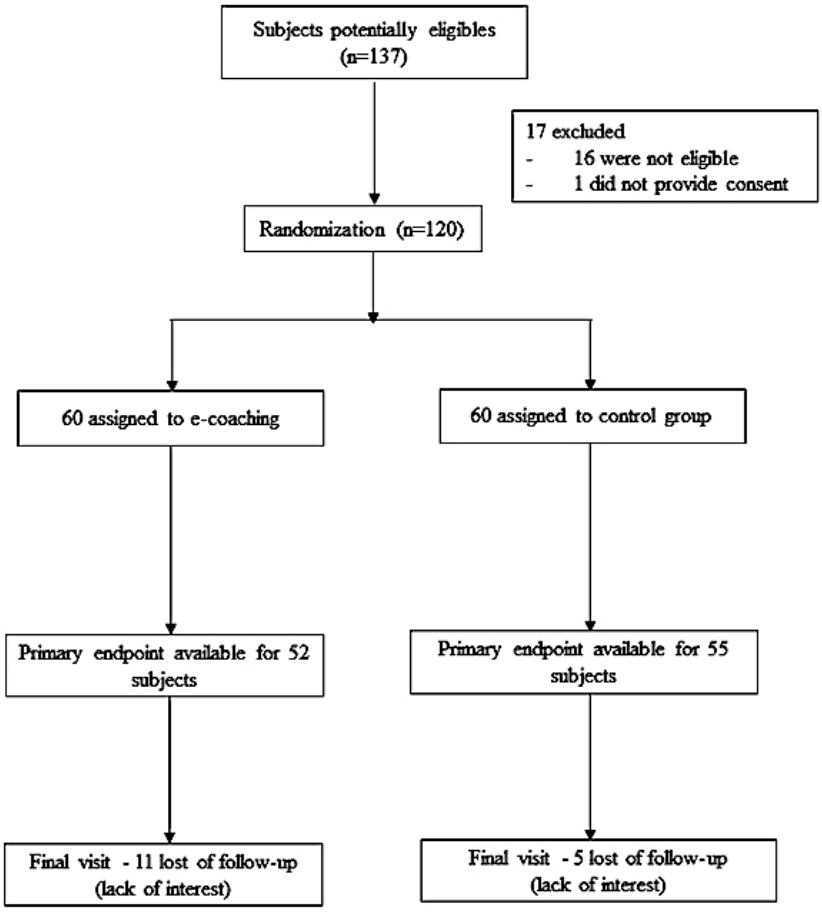


Table 2. Main primary and secondary endpoints in the intention-to-treat analysis.

Parameter	Mean difference (95% CI)		*P*-value between arms
	e-coaching (n=60)	Control (n=60)	
DQI-I score	4.55 (5.91)	-1.68 (5.18)	<.001
Energy, kcal/day	-271 (648)	-26 (663)	.11
Fibers, g/day	3.3 (12.9)	-0.7 (9.1)	.16
Carbohydrates, g/day	-24 (74)	4 (95)	.08
Proteins, g/day	-5.8 (29)	-6.8 (33)	.87
Lipids, g/day	-15.4 (34.1)	1.0 (39.3)	.02
Saturated fat, g/day	-8.6 (15.0)	3.1 (18.2)	<.001
Sodium, mg/day	-442 (1366)	175 (1981)	.07
Calcium, mg/day	-20.3 (347.3)	52.2 (428.4)	.31
Empty calories, kcal/day	-307 (663.2)	-84.6 (639.2)	.06
Body weight, kg	-2.3 (3.0)	0.2 (2.5)	.01
Waist circumference, cm	-0.9 (4.7)	0.8 (3.6)	.01
HbA1c, %	-0.30 (0.94)	0.21 (0.70)	<.001
VO_2 max, mL/minute/kg	2.99 (6.20)	2.17 (4.98)	.20

Table 3. DQI-I scores and component changes according to study arms in the intention-to-treat analysis.

Parameter	Score ranges (points)	DQI-I score between arms, mean difference (95% CI)		*P*-value between arms
		e-coaching	Control	
DQI-I score	0-100	4.55 (5.91)	-1.68 (5.18)	<.001
Variety score	0-20	0.28 (1.93)	-0.18 (1.56)	.15
Adequacy score	0-40	1.41 (4.07)	-1.00 (3.90)	.001
Moderation score	0-30	2.55 (3.43)	-0.38 (3.11)	<.001
Overall balance score	0-10	0.31 (1.04)	-0.11 (0.60)	.01
DQI-Components				
Vegetable	0-5	0.33 (1.04)	-0.28 (1.22)	.01
Fruit	0-5	0.41 (1.10)	-0.07 (1.18)	.02
Grain	0-5	0.13 (1.26)	-0.27 (1.28)	.09
Fiber	0-5	0.60 (0.96)	-0.18 (0.96)	<.001
Protein	0-5	0.03 (0.19)	0.01 (0.15)	.48
Iron	0-5	-0.14 (0.98)	-0.20 (0.90)	.75
Calcium	0-5	0.09 (0.91)	-0.06 (1.02)	.41
Vitamin C	0-5	-0.04 (1.27)	0.03 (1.47)	.76
Total fat	0-6	0.50 (0.98)	-0.15 (0.90)	<.001
Saturated fat	0-6	0.55 (1.11)	-0.16 (0.94)	.02
Cholesterol	0-6	0.62 (1.57)	-0.20 (2.12)	.02
Sodium	0-6	0.88 (1.63)	0.20 (1.70)	.03
Empty-calorie food	0-6	0.00 (0.00)	0.00 (0.00)	1

Table 4. Other secondary endpoints in the intention-to-treat analysis.

Parameter	Mean difference (95% CI)		*P*-value between arms
	e-coaching (n=60)	Control (n=60)	
Fasting blood glucose, mmol/L	-0.14 (1.46)	0.11 (1.57)	.36
Uric acid, μmol/L	-6.37 (45.45)	-8.58 (42.29)	.78
Total cholesterol, mg/dL	0.0 (0.21)	-0.04 (0.32)	.40
LDL-C, mg/dL	0.03(0.20)	-0.02 (0.28)	.43
HDL-C, mg/dL	0.00 (0.05)	0.00 (0.06)	.78
Triglycerides, mg/dL	-0.22 (1.05)	-0.14 (0.50)	.67
hs-CRP, mg/L	-0.24 (1.48)	0.08 (2.01)	.81
Serum glutamic oxaloacetic transaminase, IU/L	-3.52 (10.41)	-0.17 (12.63)	.13
Serum glutamic pyruvic transaminase, IU/L	-1.78 (6.76)	-1.28 (7.61)	.70
Gammaglutamyl-transferases, IU/L	-2.13 (20.62)	2.97 (17.73)	.47

Figure 3. Percentage of patients in the intervention and control arms achieving weight loss >3%, >5% or >10% at 16-week follow-up, in the intention-to-treat population. *P<.05 between intervention and control arms.

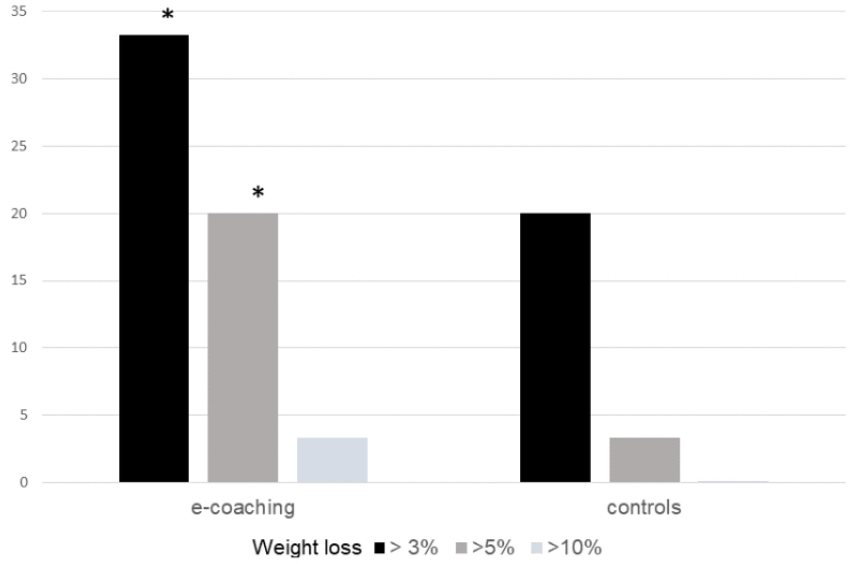


Table 5. Main primary and secondary endpoints in the per-protocol analysis

Parameter	Mean difference (95% CI)		*P*-value between arms
	e-coaching	Control	
DQI-I score[a]	5.25 (6.05)	-1.83 (5.38)	<.001
Energy, kcal/day[a]	-319 (693)	-29 (700)	.09
Fibers, g/day[a]	4 (14)	-1 (10)	.12
Carbohydrates, g/day[a]	-28 (80)	4 (101)	.08
Proteins, g/day[a]	-7 (32)	-8 (35)	.92
Lipids, g/day[a]	-18 (36)	1 (41)	.01
Saturated fat, g/day[a]	-10 (16)	3 (19)	<.001
Sodium, mg/day[a]	-521 (1470)	195 (2089)	.05
Calcium, mg/day[a]	-24 (377)	58 (452)	.32
Empty calories, kcal/day[a]	-361 (706)	-94 (674)	.05
Body weight, kg[b]	-2.9 (3.1)	0.2 (2.6)	<.001
Waist circumference, cm[c]	-1.3 (5.6)	0.90 (3.9)	.01
HbA1c, %[d]	-0.37 (1.04)	0.23 (0.73)	<.001
VO_2 max, mL/minute/kg	3.73 (6.74)	2.60 (5.36)	.09

[a]Available for 52 subjects in the intervention arm and 55 subjects in the control arm
[b]Available for 47 subjects in the intervention arm and 55 subjects in the control arm
[c]Available for 41 subjects in the intervention arm and 50 subjects in the control arm
[d]Available for 48 subjects in the intervention arm and 55 subjects in the control arm

Satisfaction Questionnaire

Of the 60 subjects of the intervention arm, 55 (92%) returned the satisfaction questionnaire by email. The overall perception of the program was good, as suggested by the comments that patients provided. Seventy percent of patients would recommend using the program for patients like them. No patient declared having encountered difficulties with the use of the platform. All the patients were able to use the platform without major difficulties and they did not need any training.

Additional Analysis

We collected the number and frequency of connections that participants made with the platform. There was a significant variation in the number of participant connections during the study. During the first month of the intervention, 93% (56/60) of participants logged in at least once per week. This percentage decreased progressively throughout the study, to reach one third of the patients in the final month.

We compared baseline characteristics of completers (subjects who completed the study) and noncompleters. The two groups were similar except for the prescription of antihypertensive drugs, which was significantly more frequent among completers (data not shown).

Interaction tests were performed to explore potential subgroups deriving a greater benefit from the intervention. No interaction was observed with respect to the primary outcome for subgroups defined according to the baseline variables (HbA1c, age, sex, BMI, education level, and occupation). In the per-protocol analysis (only including patients with complete data at the end of the study), more marked differences were observed between the ANODE arm and the control arm, and the improvement in VO_2 max tended towards statistical significance (Table 5,Table 6, and Table 7). Finally, we performed correlation analyses between the DQI-I score changes and the changes of the main variables of interest. The correlations were statistically significant for HbA1c and body weight (Table 8).

Table 6. DQI-I Scores and component changes according to study arms in the per-protocol analysis.

Parameter	Score ranges (points)	DQI-I score between arms, mean difference (95% CI)		*P*-value between arms
		e-coaching (n=52)	Control (n=55)	
DQI-I score	0-100	5.25 (6.05)	-1.83 (5.38)	<.001
Variety score	0-20	0.33 (2.07)	-0.20 (1.63)	.14
Adequacy score	0-40	1.62 (4.34)	-1.09 (4.06)	<.001
Moderation score	0-30	2.94 (3.52)	-0.42 (3.25)	<.001
Overall balance score	0-10	0.36 (1.11)	-0.12 (0.63)	.01
DQI-Components				
Vegetable	0-5	0.38 (1.41)	-0.30 (1.27)	.01
Fruit	0-5	0.47 (1.17)	-0.07 (1.23)	.02
Grain	0-5	0.15 (1.36)	-0.29 (1.34)	.09
Fiber	0-5	0.69 (1.00)	-0.19 (1.00)	<.001
Protein	0-5	0.04 (0.21)	0.01 (0.16)	.46
Iron	0-5	-0.17 (1.06)	-0.22 (0.94)	.79
Calcium	0-5	0.10 (0.97)	-0.06 (1.07)	.41
Vitamin C	0-5	-0.05 (1.36)	0.04 (1.54)	.76
Total fat	0-6	0.58 (1.17)	-0.25 (0.84)	<.01
Saturated fat	0-6	0.63 (1.17)	-0.16 (0.94)	<.001
Cholesterol	0-6	0.71 (1.67)	-0.22 (2.22)	.02
Sodium	0-6	1.02 (1.71)	0.22 (1.77)	.02
Empty-calorie food	0-6	0.00 (0.00)	0.00 (0.00)	1

Table 7. Other secondary endpoints in the per-protocol analysis.

Parameter	Mean difference (95% CI)		*P*-value between arms
	e-coaching (n=49)	Control (n=55)	
Fasting blood glucose, mmol/L	-0.18 (1.62)	0.12 (1.66)	.36
Uric acid, µmol/L	-7.80 (50.27)	-10.30 (46.22)	.80
Total cholesterol, mg/dL	0.00 (0.24)	-0.04 (0.34)	.43
LDL-C, mg/dL	0.03 (0.22)	-0.02 (0.29)	.51
HDL-C, mg/dL	0.00 (0.06)	0.00 (0.06)	.76
Triglycerides, mg/dL	-0.27 (1.16)	-0.15 (0.53)	.97
hs-CRP, mg/L	-0.31 (1.67)	0.09 (2.12)	.97
Serum glutamic oxaloacetic transaminase, IU/L	-4.31 (11.39)	-0.19 (13.32)	.12
Serum glutamic pyruvic transaminase, IU/L	-2.18 (7.43)	-1.43 (8.02)	.62
Gammaglutamyl-transferase, IU/L	-2.61 (22.84)	3.30 (18.68)	.38

Table 8. Correlation coefficients between the changes in DQI-I score and HbA1c, body weight, and waist circumference in the ANODE study.

	e-coaching	
	r^2	*P*-value
DQI-I vs HbA1c	-0.48	<.001
DQI-I score vs body weight	-0.37	<.001
DQI-I score vs waist circumference	-0.12	.21

Discussion

The randomized controlled ANODE trial in patients with T2DM and abdominal obesity showed that an automated Web-based program improved dietary habits and was associated with better weight loss and better reduction of waist circumference, compared to usual care. Moreover, glycemic control was also improved. This study is one of the first to test a fully automated but interactive Web-based program to improve lifestyle in T2DM patients. Indeed, most previous studies on telehealth for the management of T2DM have focused on telemonitoring of self-monitored blood glucose and/or were based on e-coaching conducted by phone and/or with single-way programs limited to short text messaging [4,15].

Comprehensive Analysis of Dietary Changes Associated With an Automated Web-Based Program

Several approaches to scoring dietary habits have been developed. In the present study we used the DQI-I based on 3-day dietary recall, because it measures overall diet quality [11]. To limit under- and over-reporting, self-reported data that were recorded with a validated Web-based tool [9] were verified via phone by a dietician blinded to randomization (and who did not provide any advice). The impact of an automated Web-based program on lifestyle habits, including dietetic changes, has rarely been reported in patients with T2DM. Cotterez et al [16] reviewed Internet interventions to support lifestyle modification for diabetes management. Only five such studies were identified in the literature [5,17-21]. Among them, only one concluded that dietary changes were significantly better for individuals randomized to the Web-based intervention compared to control subjects [18]. The study by Turnin et al [22], which was not included in this review, was conducted before the era of the Internet and tested a computer-assisted diet education system in patients with T2DM. This program achieved a significant improvement of dietetic knowledge and improved dietary habits; it was also associated with a decrease in HbA1c levels compared to usual care. [22]

Effect on Body Weight and Waist Circumference

A significant reduction in body weight and waist circumference was observed in the current trial. A recent review of technology-based interventions for the treatment of overweight and obese patients summarized data extracted from 27 clinical trials, and revealed the superiority of most of the interventions tested compared to control care [23]. However, unlike the ANODE program, most of these studies included time-consuming human intervention, which is an important element determining the feasibility of practical application. Two studies appeared to use fully automated programs. First, Tanaka et al [24] tested an automated tailored behavioral program. This program resulted in a significant weight loss only among obese subjects. In this subgroup, the weight loss at 3 months (-3.0 kg vs -1.4 kg in the control arm) was comparable to our results [24]. Second, in the study by Tate et al [25], participants randomized to the intervention arm who received automatically tailored messages achieved a similar weight loss at 3 months, compared to the participants randomized to the control arm who received human email counseling [25]. Taken together, these reports and the current study support the efficacy of automated interventions, at least in the intermediate term.

One question raised by the ANODE study is whether the modest weight loss obtained with e-coaching is clinically relevant. Previous modeling by the Nice Institute for Health and Care Excellence showed that at least a 1 kg weight loss among overweight or obese adults is likely to be cost-effective, provided the cost of the intervention is less than £100 and the weight difference is maintained for life [26]. The first condition is likely to be met by the program tested here because no human intervention is necessary. Whether weight loss can be maintained in the long term remains to be confirmed.

Effect on Physical Activity and Aerobic Fitness

In the ANODE study, we measured aerobic fitness by VO_2 max, which is an objective measure of aerobic activity performed over time [27,28]. Upon per-protocol analysis, the results revealed a trend towards improvement of aerobic fitness in the ANODE arm compared to the control arm. In contrast, physical activity evaluated by self-reported data was not increased in the intervention arm compared to control subjects in any of the analyses (intention-to-treat or per-protocol). Previous studies have reported conflicting data when the impact of an automated program for physical activity was assessed by questionnaires [29-32]. These questionnaires likely present certain limitations regarding capture of physical activity, especially in obese patients [33]. Another way to assess physical activity is to use trackers such as accelerometers or pedometers. Some studies have used these methods and have demonstrated the superiority of automated remote programs compared to usual care to increase physical activity [34,35]. Future studies in obese and/or patients with T2DM should objectively monitor physical activity with trackers and measure aerobic fitness in addition to self-reported data.

Limitations and Strengths

The limitations of the study include the absence of follow-up beyond 16 weeks, a relatively small sample size, and an unavoidable open-label design. In addition, only two centers were involved in recruitment. Another limitation is the lack of measurement of physical activity by trackers, although VO_2 max was measured directly. The strengths of this study include a randomized design and comprehensive analysis of dietary habits and cardiometabolic risk factors.

Feasibility of Implementing the ANODE Program

As suggested by the satisfaction questionnaire, the program was appreciated by the intervention group and was easy to use. However, because inclusion in the study required, “a frequent use (at least three times per week) of Internet as well as an email address” we cannot extrapolate this data to all patients with T2DM. Our opinion is that the feasibility of implementing such a program would be easy in a population that is accustomed to using the Internet.

Conclusion

The use of the automated Web-based ANODE e-coaching program in patients with T2DM and abdominal obesity was associated with a significant control-subtracted improvement in diet quality and several important cardiometabolic risk factors. The program can be delivered remotely with limited human resources, and therefore has potential for cost-effectiveness, and subsequently broad dissemination if generalizability and longer-term sustainability are demonstrated.

Acknowledgments

The sponsor was Assistance Publique - Hôpitaux de Paris (Département de la Recherche Clinique et du Développement; Clinical Research and Development Department). Boris Hansel is the guarantor. We acknowledge the Medical Expert Systems-MXS company (Pr. JF Boisvieux, Dr. T Boisvieux and R Boisvieux) which kindly provided the MXS-CARE e-coaching program for use in the study. The company did not take part in data analyses. We thank Karine Martin from Clinical Research Unit for study coordination and David Hajage for his participation in the statistical analyses. Assistance Publique des Hôpitaux de Paris funded the study.

Conflicts of Interest

BH was a consultant for MXS between 2009 and 2011. No other authors declare conflicts of interest.

Multimedia Appendix 1

Outlines of the e-coaching ANODE program.

[PPTX File, 1MB - jmir_v19i11e360_app1.pptx]

Multimedia Appendix 2

Baseline characteristics of the population.

[PDF File (Adobe PDF File), 26KB - jmir_v19i11e360_app2.pdf]

References

1. World Health Organization. 2016. Global report on diabetes URL: http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf [accessed 2017-09-27] [WebCite Cache ID 6tnbRd1T8]

2. Lim S, Meigs JB. Ectopic fat and cardiometabolic and vascular risk. Int J Cardiol 2013 Nov 05;169(3):166-176. [doi: 10.1016/j.ijcard.2013.08.077] [Medline: 24063931]
3. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, American Diabetes Association (ADA), European Association for the Study of Diabetes (EASD). Management of hyperglycemia in type 2 diabetes: a patient-centered approach: position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2012 Jun;35(6):1364-1379 [FREE Full text] [doi: 10.2337/dc12-0413] [Medline: 22517736]
4. Ramadas A, Quek KF, Chan CK, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
5. Noh J, Cho Y, Nam H, Kim J, Kim D, Yoo H, et al. Web-based comprehensive information system for self-management of diabetes mellitus. Diabetes Technol Ther 2010 May;12(5):333-337. [doi: 10.1089/dia.2009.0122] [Medline: 20388042]
6. Agboola S, Jethwani K, Lopez L, Searl M, O'Keefe S, Kvedar J. Text to move: a randomized controlled trial of a text-messaging program to improve physical activity behaviors in patients with type 2 diabetes mellitus. J Med Internet Res 2016 Nov 18;18(11):e307 [FREE Full text] [doi: 10.2196/jmir.6439] [Medline: 27864165]
7. Castetbon K, Vernay M, Malon A, Salanave B, Deschamps V, Roudier C, et al. Dietary intake, physical activity and nutritional status in adults: the French nutrition and health survey (ENNS, 2006-2007). Br J Nutr 2009 Sep;102(5):733-743. [doi: 10.1017/S0007114509274745] [Medline: 19250574]
8. Lassale C, Castetbon K, Laporte F, Camilleri GM, Deschamps V, Vernay M, et al. Validation of a Web-based, self-administered, non-consecutive-day dietary record tool against urinary biomarkers. Br J Nutr 2015 Mar 28;113(6):953-962. [doi: 10.1017/S0007114515000057] [Medline: 25772032]
9. Touvier M, Kesse-Guyot E, Méjean C, Pollet C, Malon A, Castetbon K, et al. Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. Br J Nutr 2011 Apr;105(7):1055-1064. [doi: 10.1017/S0007114510004617] [Medline: 21080983]
10. Lassale C, Galan P, Julia C, Fezeu L, Hercberg S, Kesse-Guyot E. Association between adherence to nutritional guidelines, the metabolic syndrome and adiposity markers in a French adult general population. PLoS One 2013;8(10):e76349 [FREE Full text] [doi: 10.1371/journal.pone.0076349] [Medline: 24124548]
11. Kim S, Haines PS, Siega-Riz AM, Popkin BM. The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. J Nutr 2003 Nov;133(11):3476-3484 [FREE Full text] [Medline: 14608061]
12. Martin A, AFSSA. Apports nutritionnels conseillés pour la population française. 3ème édition. France: Lavoisier; 2001:1-605.
13. 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]
14. WilliamsWilkins: Baltimore, MD. In: ACSM's Guidelines for Exercise Testing and Prescription. Alphen aan den Rijn, Netherlands: Wolters Kluwer; Feb 13, 2013.
15. Su D, Zhou J, Kelley MS, Michaud TL, Siahpush M, Kim J, et al. Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials. Diabetes Res Clin Pract 2016 Jun;116:136-148 [FREE Full text] [doi: 10.1016/j.diabres.2016.04.019] [Medline: 27321329]
16. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
17. Lorig K, Ritter PL, Laurent DD, Plant K, Green M, Jernigan VB, et al. Online diabetes self-management program: a randomized study. Diabetes Care 2010 Jun;33(6):1275-1281 [FREE Full text] [doi: 10.2337/dc09-2153] [Medline: 20299481]
18. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an Internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [FREE Full text] [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]
19. Carter EL, Nunlee-Bland G, Callender C. A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. Perspect Health Inf Manag 2011;8:1b [FREE Full text] [Medline: 21307985]
20. McIlhenny CV, Guzic BL, Knee DR, Wendekier CM, Demuth BR, Roberts JB. Using technology to deliver healthcare education to rural patients. Rural Remote Health 2011;11(4):1798 [FREE Full text] [Medline: 21995854]
21. McKay HG, King D, Eakin EG, Seeley JR, Glasgow RE. The diabetes network internet-based physical activity intervention: a randomized pilot study. Diabetes Care 2001 Aug;24(8):1328-1334. [Medline: 11473065]
22. Turnin MC, Bourgeois O, Cathelineau G, Leguerrier AM, Halimi S, Sandre-Banon D, et al. Multicenter randomized evaluation of a nutritional education software in obese patients. Diabetes Metab 2001 Apr;27(2 Pt 1):139-147. [Medline: 11353880]
23. Raaijmakers LC, Pouwels S, Berghuis KA, Nienhuijs SW. Technology-based interventions in the treatment of overweight and obesity: a systematic review. Appetite 2015 Dec;95:138-151. [doi: 10.1016/j.appet.2015.07.008] [Medline: 26165415]

24. Tanaka M, Adachi Y, Adachi K, Sato C. Effects of a non-face-to-face behavioral weight-control program among Japanese overweight males: a randomized controlled trial. Int J Behav Med 2010 Mar;17(1):17-24. [doi: 10.1007/s12529-009-9057-1] [Medline: 19685190]
25. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
26. National Institute for Health and Care Excellence. Public Health Draft Guideline. 2011. Maintaining a healthy weight and preventing excess weight gain among children and adults URL: https://www.nice.org.uk/guidance/ng7/documents/maintaining-a-healthy-weight-and-preventing-excess-weight-gain-among-children-and-adults-draft-guideline2 [accessed 2017-09-26] [WebCite Cache ID 6tm78CGbI]
27. Astrand PO, Rodahl K, Dahl HA, Stromme SB. Textbook of Work Physiology: Physiological Bases of Exercise, Fourth Edition. Champaign, IL: Human Kinetics; 2003.
28. Bouchard C, Blair SN, Katzmarzyk PT. Less sitting, more physical activity, or higher fitness? Mayo Clin Proc 2015 Nov;90(11):1533-1540. [doi: 10.1016/j.mayocp.2015.08.005] [Medline: 26422244]
29. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
30. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12. [doi: 10.7326/M13-3005] [Medline: 25402403]
31. Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong AT, Denney-Wilson E, et al. Effectiveness of a mHealth lifestyle program with telephone support (TXT2BFiT) to prevent unhealthy weight gain in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2015;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]
32. Jennings CA, Vandelanotte C, Caperchione CM, Mummery WK. Effectiveness of a web-based physical activity intervention for adults with type 2 diabetes-a randomised controlled trial. Prev Med 2014 Mar;60:33-40. [doi: 10.1016/j.ypmed.2013.12.011] [Medline: 24345601]
33. Rzewnicki R, Vanden AY, De BI. Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. Public Health Nutr 2003 May;6(3):299-305. [doi: 10.1079/PHN2002427] [Medline: 12740079]
34. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
35. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]

Abbreviations

ANODE: Accompagnement Nutritionnel de l'Obésité et du Diabète par E-coaching
BMI: body mass index
DQI-I: International Diet Quality Index
HbA1c: hemoglobin A1c
HDL-C: high-density lipoprotein cholesterol
hs-CRP: high-sensitivity C-reactive protein
IPAQ: International Physical Activity Questionnaire
LDL-C: low-density lipoprotein cholesterol
SD: standard deviation
T2DM: type 2 diabetes mellitus
VO2max: maximum oxygen consumption

Edited by G Eysenbach; submitted 30.04.17; peer-reviewed by K Fitzner, AS Brazeau; comments to author 26.07.17; revised version received 20.08.17; accepted 21.08.17; published 08.11.17.

Please cite as:
Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, Kadouch D, Hartemann A, Oppert JM, Bruckert E, Marre M, Bruneel A, Duchene E, Roussel R
A Fully Automated Web-Based Program Improves Lifestyle Habits and HbA1c in Patients With Type 2 Diabetes and Abdominal Obesity: Randomized Trial of Patient E-Coaching Nutritional Support (The ANODE Study)
J Med Internet Res 2017;19(11):e360
URL: http://www.jmir.org/2017/11/e360/
doi:10.2196/jmir.7947
PMID:29117929

Original Paper

Brain Activation in Response to Personalized Behavioral and Physiological Feedback From Self-Monitoring Technology: Pilot Study

Maxine E Whelan[1,2], BSc; Paul S Morgan[1,2,3], PhD; Lauren B Sherar[1,2,4], PhD; Andrew P Kingsnorth[1,2], PhD; Daniele Magistro[1,2], PhD; Dale W Esliger[1,2,4], PhD

[1]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom
[2]National Centre for Sport and Exercise Medicine, Loughborough University, Loughborough, United Kingdom
[3]Medical Physics and Clinical Engineering, Nottingham University Hospitals, Nottingham, United Kingdom
[4]National Institute of Health Research Leicester Biomedical Research Centre, Leicester, United Kingdom

Corresponding Author:
Maxine E Whelan, BSc
Loughborough University
School of Sport, Exercise and Health Sciences
Epinal Way
Loughborough,
United Kingdom
Phone: 44 1509226302
Email: m.e.whelan@lboro.ac.uk

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2017/12/e430/

Abstract

Background: The recent surge in commercially available wearable technology has allowed real-time self-monitoring of behavior (eg, physical activity) and physiology (eg, glucose levels). However, there is limited neuroimaging work (ie, functional magnetic resonance imaging [fMRI]) to identify how people's brains respond to receiving this personalized health feedback and how this impacts subsequent behavior.

Objective: Identify regions of the brain activated and examine associations between activation and behavior.

Methods: This was a pilot study to assess physical activity, sedentary time, and glucose levels over 14 days in 33 adults (aged 30 to 60 years). Extracted accelerometry, inclinometry, and interstitial glucose data informed the construction of personalized feedback messages (eg, average number of steps per day). These messages were subsequently presented visually to participants during fMRI. Participant physical activity levels and sedentary time were assessed again for 8 days following exposure to this personalized feedback.

Results: Independent tests identified significant activations within the prefrontal cortex in response to glucose feedback compared with behavioral feedback (P<.001). Reductions in mean sedentary time (589.0 vs 560.0 minutes per day, P=.014) were observed. Activation in the subgyral area had a moderate correlation with minutes of moderate-to-vigorous physical activity (r=0.392, P=.043).

Conclusion: Presenting personalized glucose feedback resulted in significantly more brain activation when compared with behavior. Participants reduced time spent sedentary at follow-up. Research on deploying behavioral and physiological feedback warrants further investigation.

(J Med Internet Res 2017;19(11):e384) doi:10.2196/jmir.8890

KEYWORDS

functional magnetic resonance imaging; neuroimaging; physical activity; sedentary behavior; interstitial glucose

Introduction

Physical inactivity, insufficient levels of physical activity, is attributable to 9% of premature mortality and 7% of type 2 diabetes cases [1]. In addition, sedentary behavior, defined as "any waking behavior characterized by an energy expenditure ≤1.5 metabolic equivalents of task (METs) while in a sitting or reclining posture" [2], has been strongly associated with poor cardiometabolic health [3]. With adults spending an estimated 7 hours sedentary each day [4] and the prevalence of type 2 diabetes expected to rise to 592 million by 2035 [5], it is critical to address the prevalence of physical inactivity and time spent sedentary for the amelioration of type 2 diabetes and other important chronic, noncommunicable diseases.

Over the last decade, wearable activity monitors have grown in popularity in consumer markets to help users physically track their movement behaviors (eg, active minutes, step counts, distance traveled, time spent sitting) [6,7]. Over this same time, wearable physiological sensing devices (eg, heart rate monitors, continuous glucose monitors) have been evolving and are now venturing beyond the clinical domains and into more consumer-focused markets [8]. The allure of these wearable technologies is that they provide users with real-time personalized health feedback that may act to encourage positive lifestyle behaviors (eg, moving more, sitting less, eating more healthily) [9]. However, with 32% of individuals failing to continue using these devices beyond 6 months following purchase [10], there is a need to optimize the feedback provided to the users to maintain adoption and sustain engagement with the information presented. Patel and colleagues [11] suggest that providing explanatory feedback in an understandable manner is important to encourage sustained use. Given that sustained behavior change is often poorly reported and not often achieved [12], assessing how people respond to this feedback at a cortical level (by monitoring changes in brain activation) could reveal additional insight above traditional routes such as focus groups or interviews.

Neuroimaging techniques are useful to recognize and identify the intricate relationships between cognitions, brain functions, and behavior [13]. There has been growing interest in the community toward communication neuroscience, research that provides a deep understanding of attitude and behavior change [14]. Moreover, communication neuroscience research suggests that people's intentions and behavior are largely affected by the content and format of an advertisement [15]. One key neuroimaging tool is functional magnetic resonance imaging (fMRI), which can monitor neural responses as information is presented [16] (eg, health messages and advertisements [14,17,18]). Receiving personalized (or self-related) feedback is often associated with activation within the rostral medial prefrontal cortex (mPFC), associated with decision making and mimicry behavior [19,20], and the precuneus/posterior cingulate region, often associated with personal reflection [21-23]. In particular, self-relevant messages elucidate more activation within the mPFC than nontailored messages [24] and can predict behavior change [25]. Meta-analyses of functional neuroimaging studies also suggest that the mPFC and precuneus/posterior cingulate regions mediate self-related processing [26,27].

fMRI can improve our understanding of how cognitive processes vary between those who do change their behavior following exposure to a stimulus and those who do not subsequently change [28]. The mPFC is positioned whereby activation in this region can predict individual behavior change [14,17,29]. To date, research has largely focused on identifying neural responses to antismoking material [17,29,30] rather than diet, alcohol consumption, physical inactivity, or sedentary behavior [31]. Investigating how people respond to personalized feedback relating to these lifestyle behaviors could offer crucial insight into how best to disseminate feedback to maximize effect and thus help to design materials that optimize population health [32]. For instance, observed reductions in smoking rates have been attributed to a number of influences, in part, by the dissemination of antismoking materials (eg, cigarette packaging labels) [33]. Given that the literature to date has largely assessed how people respond to antismoking materials, fMRI may help identify how people's brains respond to information commonly presented on the screens of wearable devices and associated smartphone apps. The authors hypothesize that the mPFC and precuneus/posterior cingulate regions will be activated given the presentation of personalized or self-relevant feedback [21-23,26,27]. The aims of this study were to identify regions of the brain activated in response to personalized behavioral and physiology feedback messages and examine behavior change and associations with levels of brain activation.

Methods

Participants

A total of 33 participants (57% female) were recruited from a university in the United Kingdom via advertisement posters and email. Participants were aged 30 to 60 years, had no mobility-related musculoskeletal problems, had no confirmed diagnosis of diabetes, were willing and able to comply with the study protocol, met standard fMRI safety criteria (no metal in body, not claustrophobic, not pregnant), and were right-handed. All participants completed a physical activity readiness questionnaire [34] prior to participation with positive responses assessed by a clinician.

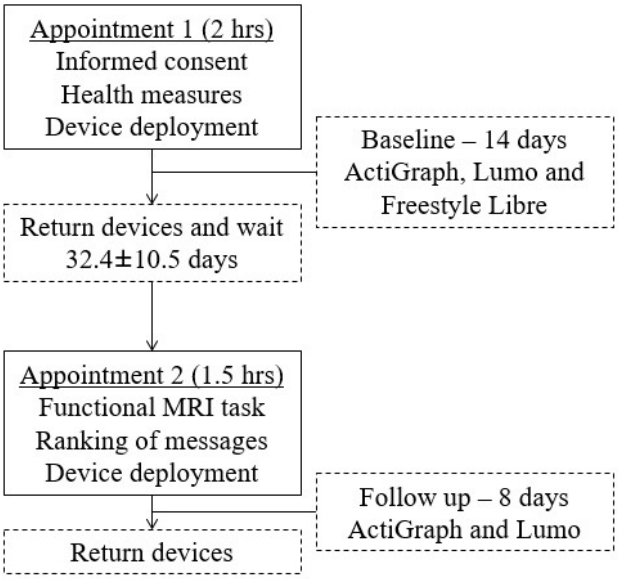
Each participant's consent was obtained according to the Declaration of Helsinki, and all experimental procedures were approved by the Loughborough University Ethics Advisory Committee (R15-P142).

Procedure

Data were collected between June and September 2016. The study design is presented in Figure 1. During the first appointment, participants provided informed consent; answered questions relating to age, gender, ethnicity, and education; and completed a selection of health measures (body composition, blood pressure, and blood sample). Participants were fitted with 3 devices to monitor their physical activity, sedentary behavior, and glucose levels for 14 days. In addition, participants were provided an education booklet to read prior to the fMRI appointment. This booklet included background information about physical activity (eg, moderate-to-vigorous intensity [MVPA]) and also offered recommendations (eg, target range for glucose levels) to help minimize any variations in

knowledge. The fMRI took place at the second appointment (on average 32.4 [SD 10.5] days following the first appointment); following this, participants continued to wear 2 devices to monitor physical activity and sedentary behavior for 8 days. At the end of the follow-up period, participants returned the devices and received a comprehensive personalized health report.

Figure 1. Study design.



Measures

Physical Health

Weight and body fat percentage were measured using the MC 780 MA scale (Tanita) following the removal of shoes and socks. Body mass index was calculated as weight (kg) divided by height (m) squared (weight/height^2). Glucose and hemoglobin A_{1c} (HbA_{1c}) were analyzed using a Cholestech LDX system and Afinion AS100 Analyzer (both Alere Inc), respectively. Participants arrived fasted for ≥8 hours prior to the collection of a capillary blood sample.

Accelerometry

An ActiGraph wGT3x-BT accelerometer (ActiGraph LLC) was worn on a waistband (on the right anterior axillary line) to objectively measure physical activity. Participants were asked to wear the validated device [35] during waking hours and to remove for any water-based activities (eg, showers or bathing). The accelerometry data were collected at 100 Hz resolution and integrated into 60 second epochs using ActiLife version 6.13.2 (ActiGraph LLC) and processed using Kinesoft version 3.3.80 (Kinesoft). Sedentary behavior, light activity, and MVPA were defined as ≤100 counts per min (cpm), 101 to 2019 cpm, and >2019 cpm, respectively [36]. Nonwear was identified by an interval of at least 60 consecutive minutes of zero activity intensity counts, with allowance for 1 to 2 minutes of counts between 0 and 100 [36]. Participants who had <4 valid days were excluded from analyses. A valid day was defined as having ≥10 hours of monitor wear. Accelerometers were initialized to begin monitoring at the end of appointments, which meant participants had a variable amount of possible wear on the first day. As a result, to standardize the opportunity for participants to adhere to device wear, days 2 through 8 were analyzed for both baseline and follow-up. A global wear time variable was calculated as the mean of wear time at baseline and follow-up.

Inclinometry

A Lumo (Lumo Bodytech Inc) posture sensor was worn on a waistband (in the lumbosacral region) in contact with the skin to measure sedentary behavior (time spent sitting, driving, lying, standing, stepping, and number of sit-to-stand transitions) during baseline and follow-up. Devices were calibrated to the wearer. Participants were asked to wear the device only during waking hours, remove it for any water-based activities (eg, showers or bathing), and place the device on charge overnight each day. The Lumo has been found to produce valid measurements of sedentary behavior compared with the ActivPAL (PAL Technologies Ltd), with a mean error of 9.5% [37]. Data from the Lumo devices were analyzed in 5-minute epochs (highest resolution) using Excel (Microsoft Corp). Nonwear was defined by 1 of 2 criteria: (1) device removal for sleep which was automatically detected if the device was placed on charge or (2) prolonged periods of the same posture deemed to be biologically implausible (ie, ≥60 minutes). Again, the Lumos were set up to begin monitoring at the end of appointments, and days 2 through 8 were analyzed for both baseline and follow-up.

Flash Glucose Monitoring

The Freestyle Libre flash glucose monitor (Abbott Laboratories) measures interstitial glucose levels via a minimally invasive 5 mm flexible filament inserted into the posterior upper arm. The sensor works based on the glucose-oxidase process by measuring an electrical current proportional to the concentration of glucose. Tegaderm transparent film dressing (3M Health Care) was applied on top of the sensor to maintain its position. Participants were informed not to remove the sensor and to scan at least once every 7 hours (a conservative decision as the manufacturer

states 8 hours to avoid data loss). As a result, participants were able to see their real-time glucose levels during baseline wear. An indication of how many times participants viewed this information (level of exposure) was identified by the number of time they scanned. Missing data were obtained because of a fault (sensor lasted <14 days) or the participant failed to scan at least once every 8 hours. The Freestyle Libre has been previously validated against venous sampling with an overall mean absolute relative difference of 11.4% with consistent accuracy throughout the 14 days [38]. Glucose data were downloaded in 15-minute epochs (highest resolution) using Freestyle Libre version 1.0. The raw data were used to calculate the number of high glucose events (defined as ≥8.8 mmol/L) and identify valid days. Days were defined as valid if they met the prespecified threshold of ≥90% of data points (96 expected based on 4 readings each hour across each 24-hour period). All 14 days were analyzed from baseline wear. Area under the curve was calculated from the mean area of the positive peaks across the valid days using GraphPad Prism version 7.0.0 (GraphPad Software), and participants' fasting glucose levels were used as the baseline.

Functional Magnetic Resonance Imaging Stimuli

Twenty personalized feedback messages were created for the purposes of this study and covered 4 topics: MVPA, light physical activity, sedentary behavior, and glucose levels (all presented in Figure 2 with example data). They intended to reflect feedback metrics commonly presented on wearable technologies. Data obtained via accelerometry, inclinometry, and flash glucose monitoring were analyzed and then incorporated into the personalized feedback messages. Therefore, the values presented on the messages were personalized so that the numbers varied from 1 participant to another but the image and text remained consistent. The images were matched in visual complexity, color, and text font using Axure RP Pro version 7.0.0.3190 (Axure Software Solutions Inc) to standardize the stimuli across participants. Picture icons were identified and downloaded from an icon resource website (www.flaticon.com).

Stimuli were presented on a monitor located 2.8 m behind the center of the scanner bore and viewed by a mirror mounted on the head coil. Adjustments to the positioning of the mirror were made for participants to ensure that the full monitor screen could be seen. We examined neural activity while participants were presented with feedback and were requested to maintain attention throughout. Prior to the start of the fMRI task, there was an initial period of 40 seconds of dummy scans which were immediately discarded. The fMRI task is outlined in Figure 3. In total, 24 blocks (12 active, 12 rest) were presented during the protocol. Each active block consisted of stimulus presentation of 5 back-to-back trials (referred to as images from this point forward) of 8 seconds each, totaling 40 seconds, followed by a rest period of 40 seconds, during which participants viewed a fixation cross and were instructed to clear their minds. The blocks and back-to-back images (within the blocks) were not presented in a counterbalanced or randomized order.

Figure 2. Personalized feedback stimuli.

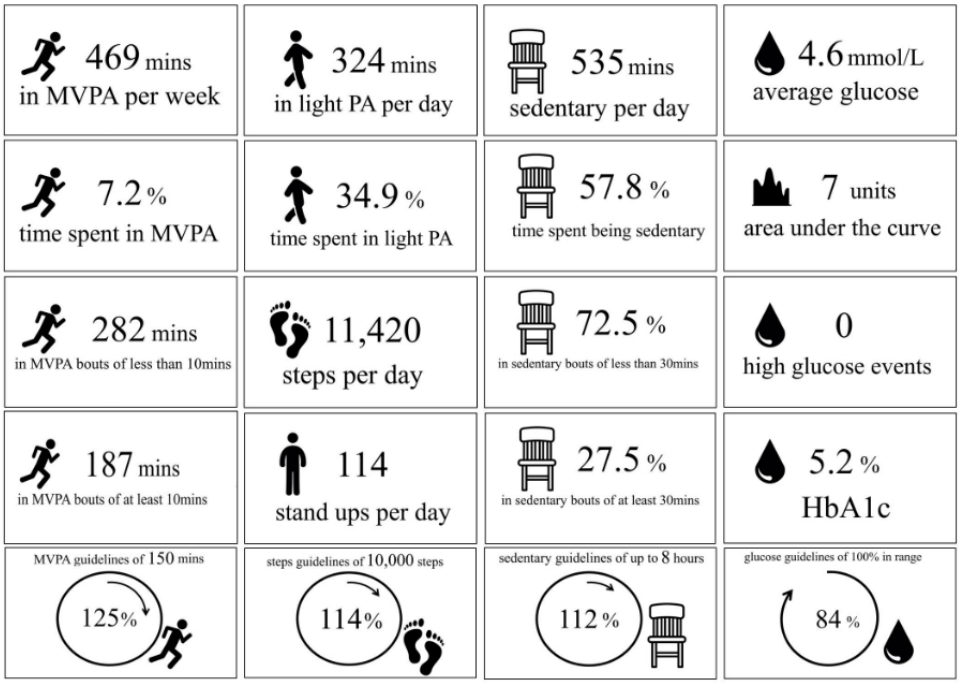
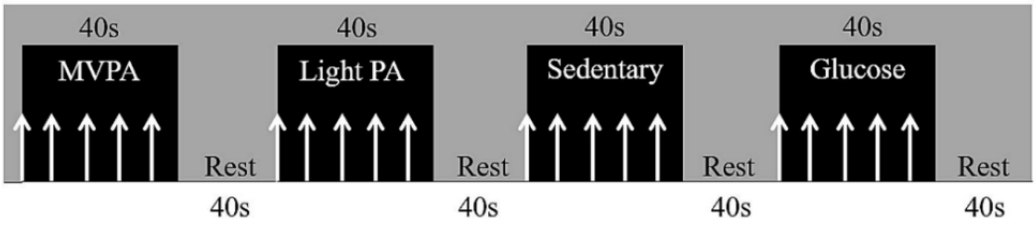


Figure 3. Trial setup including 8 of the 24 blocks presented.



Functional Magnetic Resonance Imaging Data Acquisition

Brain imaging data were acquired on a 3T Discovery MR750w scanner (General Electric) using a 32-channel head coil at the National Centre for Sport and Exercise Medicine, Loughborough University, United Kingdom. Structural images (T1-weighted) were acquired using a fast spoiled gradient echo (FSPGR) Bravo sequence (3D volume, FSPGR; TR=8.2 ms; TE=3.1 ms; matrix size 240×240; 160 sagittal slices; FOV=240 mm; 1 mm thick). One functional scan lasting 16 minutes (480 volumes) was acquired during the task (2D gradient echo EPI; TR=2000 ms; TE=30 ms; flip angle=75 degrees; matrix size 64×64; 35 axial slices; FOV=205 mm; 3 mm thick). Stimulus presentation and synchronization to scanner acquisition were performed using Presentation version 18.1 (Neurobehavioral Systems Inc).

Analyses

Functional Magnetic Resonance Imaging Data Analysis

Functional data were preprocessed and analyzed using statistical parametric mapping (SPM12, Wellcome Department of Cognitive Neurology). All data reported are from scans that exhibited ≤3 mm in translational movement. Data were processed using a standard statistical parametric mapping approach, which consisted of scan realignment, coregistration, segmentation, normalization, and smoothing. Data were spatially aligned to the first functional image using 4th degree B-spline interpolation. Scans were then coregistered (mean functional image aligned with T1 then parameters applied to all functional images). Functional images were normalized into the Montreal Neurological Institute (MNI) standard stereotactic space with parameters applied to all functional images. A final smoothing step with a Gaussian Kernel with full width half maximum of 8 mm was applied to improve signal-to-noise ratio. The onsets and durations of each of the conditions of interest were modeled according to the block design described in the protocol. For each participant, brain activation was estimated using a general linear model (GLM) and included movement parameters (3 translations, 3 rotations) and a session constant as regressors. All regressors were convolved with SPM12's canonical difference of the hemodynamic response function. Data were high-pass filtered with a cut off of 128 seconds to remove low-frequency noise and slow drifts in the signal. Family-wise error (FWE) correction was used to correct for multiple comparisons at P_{FWE}. At the first level for each participant, contrasts were computed using a series of univariate analyses of covariance (ANCOVAs), averaging activity across the topics compared with baseline: (1) MVPA>baseline, (2) light physical activity>baseline, (3) sedentary>baseline, (4) glucose>baseline, and (5) behavior>baseline. Additional contrasts were computed using a series of univariate ANCOVAs, averaging activity between the different blocks of stimuli (ie, MVPA>light physical activity and glucose>sedentary) and reverse contrasts also computed (ie, light physical activity>MVPA and sedentary>glucose).

Second level random effects models for each task were constructed that averaged across participants and were subjected to further region of interest (ROI) and between-group analysis (described below). Exploratory whole brain searches were conducted for each contrast with a threshold set at $P<.001$ and $P<.05$ for the baseline contrasts and intergroup contrasts, respectively (cluster threshold of k=0 voxels). Between-group analyses were conducted to compare gender differences and differences between those least (<150 minutes of MVPA per week) and most (≥150 minutes of MVPA per week) active. Using independent samples *t* test analysis, brain regions were labeled according to the MNI anatomic labeling tool implemented in the Wake Forest University Pickatlas (WFU Pickatlas) [39]. The average beta parameter estimates of activity during the presentation of information compared with other information blocks were extracted using MarsBaR, an ROI toolbox. All models controlled for centered demographic variables (centered age and sex). An additional centered variable (number of daily glucose scans) was included within the additional contrasts conducted.

Statistical Analysis

To examine demographic and self-report data, we conducted descriptive analyses using SPSS version 22.0 (IBM Corp). Two group *t* tests were conducted to produce descriptive outcomes. Repeated measures ANCOVAs were conducted to assess changes in behavior (levels of MVPA, light physical activity, and sedentary behavior) from baseline to follow-up, controlling for global wear time (average wear time). Tests of statistical significance were based on 2-sided probability ($P<.05$).

Correlation Analysis

Parameter estimates corresponding to each significantly activated region, identified via fMRI data analysis, were extracted for each participant. Linear regressions provided partial correlation coefficients between the parameter estimates from the significant regions of interest and subsequent behavior at follow-up (ie, time spent in MVPA, light physical activity, and sedentary), controlling for wear time. The relationships between behavior change and activity from the ROIs were examined in separate models for each ROI, and the analyses were repeated to assess behavior via both accelerometry and inclinometry data.

Results

Participants

A flow chart of individuals through the study and the characteristics of the study sample are presented in Figure 4 and Table 1, respectively. Four participants were excluded from fMRI analyses due to incorrect scanner parameter setup, poor participant vision (without glasses), and presence of an unsafe magnetic resonance implant. One participant fell asleep, and an additional participant was excluded due to incorrect accelerometry initialization. This resulted in a final sample of 28 participants for the full study protocol.

Figure 4. Flowchart of individuals at each stage of the study.

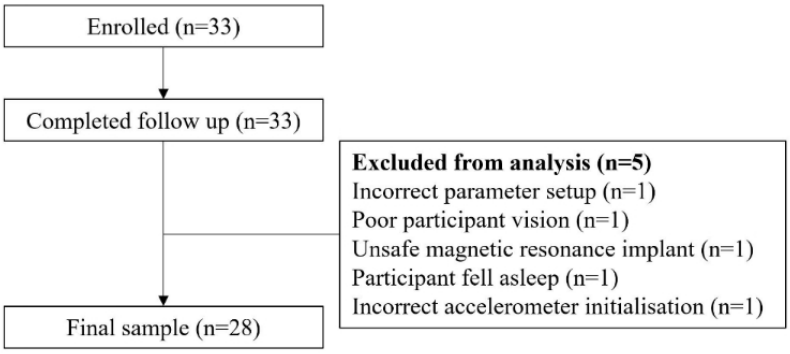


Table 1. Sample characteristics.

Characteristics	Whole sample (n=28)
Demographic	
Age (years), mean (SD)	44.2 (9.5)
Male, %	42.9
Baseline	
Weight (kg), mean (SD)	75.2 (15.3)
Body mass index (kg/m^2), mean (SD)	25.2 (4.3)
Body fat (%), mean (SD)	26.7 (9.3)
HbA_{1c}[a] (%), mean (SD)	5.4 (0.4)
Glucose (mmol/L), mean (SD)	5.0 (0.6)

[a]HbA_{1c}: hemoglobin A_{1c}.

The 28 participants (43% male) had a mean age of 44.2 (SD 9.5) years (range 30 to 59 years). Three (11%) participants completed secondary school, 5 (18%) completed some additional training, and 20 (71%) received a bachelor's degree or higher. Twenty-five (89%) were white, 2 (7%) were Chinese, and 1 (4%) was Asian or Asian British. Males were significantly taller (178.7 versus 167.5 cm), had a lower body fat percentage (18.8% versus 32.6%), and scanned the Freestyle Libre more frequently (9.5 versus 5.7 scans per day).

Activated Regions of the Brain

First, we contrasted each of the 4 topics with a fixation cross. The brain regions significantly activated in response to the initial contrasts of interest are presented in Table 2. Regions include the middle and inferior occipital gyrus, middle frontal gyrus, lingual gyrus, subgyral, and thalamus (P<.001). No significant voxels were identified between those most and least active or between males and females.

We then proceeded to the main analysis that contrasted the topics between themselves. The brain regions identified as significantly activated are presented in Table 3. Of the additional contrasts of interest, the glucose>behavior contrast highlighted significant activation in the middle frontal gyrus (–32, 36, –12, z=5.60) and left subgyral (–26, 48, 4, z=5.33). The glucose>sedentary contrast revealed significant activation in the cuneus (–2, –80, 4, z=5.05), middle frontal gyrus (–32, 36, –12, z=4.95; –20, 34, 42, z=4.94), superior frontal gyrus (–26, 50, 4, z=4.79), and right subgyral (28, –52, 24, z=4.66) (Figure 5, Table 3).

Table 2. Average contrasting differences (thresholded at *P*<.001, cluster threshold of k=0 voxels).

Region	MNI[a] coordinates							
	Hem[b]	x	y	z	Voxels	Z	t	P_{FWE} [c]
MVPA[d]**>baseline**								
Middle occipital gyrus	L	–38	–74	–14	178	6.29	9.99	<.001
Lingual gyrus	L	–14	–94	–10	—	6.25	9.89	<.001
Inferior occipital gyrus	L	–22	–90	–14	—	6.21	9.76	<.001
Subgyral	R	36	–62	–16	11	6.06	9.29	<.001
Fusiform gyrus	L	–36	–54	–16	9	5.97	9.03	<.001
Subgyral	R	34	–84	–6	93	5.95	8.97	<.001
Lingual gyrus	R	24	–92	–10	—	5.86	8.74	<.001
Lingual gyrus	R	16	–90	–8	—	5.63	8.11	.001
Inferior occipital gyrus	R	44	–76	–12	2	5.62	8.09	.001
Middle occipital gyrus	R	30	–88	4	1	5.57	7.97	.001
Light PA [e] **>baseline**								
Cuneus	L	–16	–96	–2	101	6.47	10.61	<.001
Middle occipital gyrus	L	–32	–84	–14	119	6.23	9.80	<.001
Middle occipital gyrus	L	–38	–72	–14	—	6.05	9.28	<.001
Subgyral	R	34	–84	–6	83	6.05	9.26	<.001
Middle occipital gyrus	R	30	–84	–14	—	5.68	8.24	.001
Middle occipital gyrus	R	46	–76	–10	23	6.01	9.14	<.001
Subgyral	R	36	–62	–16	23	5.90	8.83	<.001
Middle occipital gyrus	R	28	–98	6	19	5.77	8.48	<.001
Fusiform gyrus	L	–36	–54	–16	3	5.77	8.47	<.001
Fusiform gyrus	L	–34	–50	–18	2	5.70	8.30	<.001
Inferior frontal gyrus	L	–54	18	20	4	5.69	8.27	<.001
Lingual gyrus	R	16	–90	–10	10	5.65	8.18	.001
Sedentary>baseline								
Middle occipital gyrus	L	–36	–72	–16	19	5.99	9.11	<.001
Inferior occipital gyrus	L	–38	–82	–10	46	5.95	8.98	<.001
Subgyral	L	–20	–94	–6	36	5.87	8.77	<.001
Middle occipital gyrus	R	36	–84	–4	4	5.78	8.50	<.001
Inferior frontal gyrus	L	–48	14	22	3	5.65	8.16	.001
Middle occipital gyrus	R	48	–76	–10	3	5.59	8.02	.001
Subgyral	R	28	–88	–6	1	5.57	7.97	.001
Glucose>baseline								
Cuneus	L	–16	–96	–6	218	6.69	11.38	<.001
Middle occipital gyrus	L	–36	–74	–16	—	6.13	9.50	<.001
Middle occipital gyrus	L	–20	–90	–14	—	5.90	8.83	<.001
Subgyral	R	36	–62	–16	13	5.99	9.10	<.001
Lingual gyral	R	14	–90	–8	28	5.97	9.05	<.001
Subgyral	R	28	–84	–6	56	5.88	8.78	<.001
Middle frontal gyrus	L	–40	10	30	6	5.69	8.27	<.001

Region	MNI[a] coordinates							
	Hem[b]	x	y	z	Voxels	Z	t	P_{FWE}[c]
Middle occipital gyrus	R	44	–76	–14	1	5.60	8.05	.001
Middle occipital gyrus	R	30	–84	–14	2	5.58	8.00	.001
Behavior>baseline								
Middle occipital gyrus	L	–38	–72	–16	272	6.44	10.49	<.001
Cuneus	L	–16	–96	–6	—	6.33	10.12	<.001
Middle occipital gyrus	L	–32	–84	–14	—	6.07	9.33	<.001
Subgyral	R	36	–62	–16	27	6.16	9.61	<.001
Subgyral	R	34	–84	–6	135	6.14	9.53	<.001
Lingual gyral	R	22	–92	–10	—	5.85	8.69	<.001
Middle occipital gyrus	R	30	–84	–14	—	5.75	8.42	<.001
Superior parietal lobule	L	–32	–62	58	5	6.06	9.28	<.001
Middle occipital gyrus	R	46	–76	–12	24	5.96	9.00	<.001
Middle occipital gyrus	R	48	–66	–14	—	5.88	8.79	<.001
Fusiform gyrus	L	–36	–54	–16	8	5.95	8.98	<.001
Middle frontal gyrus	L	–52	26	26	9	5.73	8.38	<.001
Thalamus	R	22	–28	–2	2	5.69	8.27	.001

[a]MNI: Montreal Neurological Institute.

[b]hem: hemisphere.

[c]FWE: family-wise error.

[d]MVPA: moderate-to-vigorous physical activity.

Table 3. Average contrasting differences controlling for age, gender, and average daily number of glucose scans (thresholded at $P<.05$, cluster threshold of k=0 voxels).

Region	MNI[a] coordinates							
	Hem[b]	x	y	z	Voxels	Z	t	P_{FWE}[c]
Glucose>behavior								
Middle frontalgyrus	L	–32	36	–12	25	5.60	8.17	<.001
Subgyral	L	–26	48	4	16	5.33	7.48	<.001
Glucose>sedentary								
Cuneus	L	–2	–80	4	34	5.05	6.85	<.001
Middle frontal gyrus	L	–32	36	–12	8	4.95	6.63	<.001
Middle frontal gyrus	L	–20	34	42	11	4.94	6.61	<.001
Superior frontal gyrus	L	–26	50	4	3	4.79	6.29	<.001
Subgyral	R	28	–52	24	1	4.66	6.04	<.001

[a]MNI: Montreal Neurological Institute.

[b]Hem: hemisphere.

[c]FWE: family-wise error.

Figure 5. Group level significant activation pattern for the contrast glucose>behavior at the MNI coordinates (a) –32, 36, –12 and (b) –26, 48, 4.

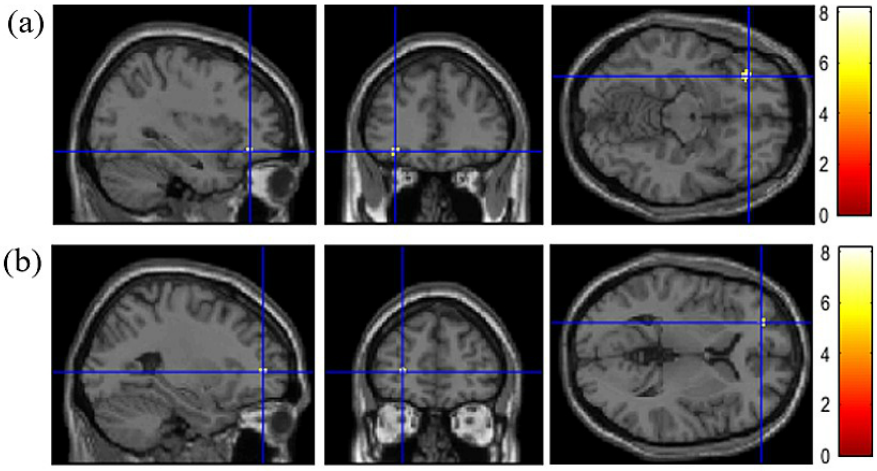


Table 4. Behavioral characteristics derived from accelerometry and inclinometry.

	Accelerometry[a], mean (SD)			Inclinometry[b], mean (SD)		
	Baseline	Follow-up	*P* value	Baseline	Follow-up	*P* value
Number of valid days	7.0 (0.0)	7.0 (1.0)	—	4.2 (2.1)	5.5 (1.7)	—
Wear time	903.5 (67.7)	868.2 (70.4)	.002	924.3 (61.9)	884.0 (61.6)	.001
Step count	9065.2 (3456.2)	9634.0 (3699.3)	—	8660.9 (2995.7)	9580.3 (4326.0)	—
Counts per minute	194.0 (82.0	410.0	<.001	—	—	—
Sedentary (min)	589.0 (84.7)	560.0 (75.6)	.014	602.2 (91.1)	554.5 (89.4)	.001
Light PA[c] (min)	265.0 (69.0)	254.2 (71.1)	—	—	—	—
Moderate (min)	45.8 (31.0)	50.7 (33.2)	—	—	—	—
Vigorous (min)	3.6 (6.6)	3.2 (6.2)	—	—	—	—
MVPA[d] (min)	49.4 (34.2)	53.9 (35.5)	—	—	—	—
LVPA[e] (min)	314.4 (66.4)	308.1 (72.1)	—	—	—	—
Stepping (min)	—	—	—	93.5 (26.7)	103.2 (44.1)	—
Standing (min)	—	—	—	228.5 (98.5)	226.5 (67.8)	—

[a]≥4 valid days, n=28 (100% compliance to ≥600 mins of accelerometer wear).

[b]≥1 valid day, n=23, (100% compliance to ≥600 mins of inclinometry wear).

[c]PA: physical activity.

[d]MVPA: moderate-to-vigorous physical activity.

[e]LVPA: light-to-vigorous physical activity.

Behavior Change

The behavioral characteristics obtained via accelerometry and inclinometry are presented in Table 4. Among the 28 participants, 100% provided ≥4 days for accelerometry during baseline and follow-up. In contrast, only 15 (54%) and 20 (71%) participants provided ≥4 valid days at baseline and follow-up with the inclinometer, respectively, revealing a reduced sample (13 vs 28). As a result, the criteria for the Lumo was adjusted to ≥1 valid days. From baseline to follow-up, wear time and sedentary time reduced while minutes of MVPA and counts per minute increased. After controlling for global wear time, only time spent sedentary remained significant for both the accelerometry and inclinometry (589.0 [SD 13.9] minutes vs 560.0 [SD 11.7] minutes, *P*=.014 and 602.2 [SD 19.4] minutes vs 554.5 [SD 18.1] minutes, *P*=.001, respectively). Despite a lack of change at the whole sample level for time spent in light physical activity, MVPA, and step count, 9 (32%), 17 (61%), and 16 (57%) participants, respectively, positively increased the amount of steps, light physical activity, and MVPA at follow-up (unadjusted for global wear time). Males accumulated significantly more vigorous physical activity compared with females at baseline and follow-up (*P*=.029 and *P*=.026, respectively) and also significantly more minutes of MVPA (*P*=.033) at follow-up. No significant associations were observed between number of scans and changes in behavior via accelerometry or inclinometry (controlling for global wear time).

Functional Magnetic Resonance Imaging Correlations

To investigate the relationship between brain activation and subsequent behavior, parameter estimates were calculated for the patterns of neural activation. Of these, only glucose feedback was positively associated with subsequent minutes of MVPA (r=0.392, P=.043). No significant associations were observed for the inclinometry data.

Discussion

Summary

As recent neuroimaging work has highlighted value in analyzing individual responses to feedback relating to lifestyle behaviors [14], we used fMRI to examine neural responses to personalized feedback relating to physical activity, sedentary behavior, and interstitial glucose levels. We also investigated associations between neural activity and subsequent behavior. This study lies at the intersection of 3 rapidly evolving areas of interest: wearables, lifestyle behaviors, and neuroimaging. Our study identified that presenting people with personalized feedback relating to interstitial glucose levels resulted in significantly more brain activation when compared with personalized behavioral feedback.

Activated Regions of the Brain

Our findings identified activations within regions of the prefrontal cortex, in particular the middle frontal gyrus, subgyral, cuneus, and superior frontal gyrus upon comparison of personalized glucose feedback with behavioral feedback. Previous studies have also identified regions within the prefrontal cortex following exposure to antismoking images [29], messages encouraging sunscreen use [14], and informative nutritional labels [40]. The authors hypothesized that the mPFC and precuneus/posterior cingulate regions would be activated in our study given the presentation of personalized and self-relevant feedback [21-23,26,27]. Other fMRI studies have identified alternate activated regions including the ventromedial prefrontal cortex, inferior frontal gyrus, and amygdala when presented information about other lifestyle behaviors (eg, smoking) [30,41,42]. The findings suggest that the personalized feedback did not offer identical regions of interest when compared with the literature; however, some activation did overlap with the mPFC. Neuroimaging studies impose additional complexity because identical neural patterns can result after exposure to different stimuli [43]. However, the identified regions of brain activation may also differ because the stimuli differs between fMRI studies. Our study used a combination of text and images to inform participants about their physical activity, sedentary behavior, and interstitial glucose levels. In comparison, Falk and colleagues [41] presented images with text and numbers in a sentence (multiple lines of text) format. Overall, our findings suggest that it is possible to identify what brain regions are activated in response to personalized feedback and that glucose-related feedback evoked more brain activation. As a result, wearable technologies presenting personalized glucose feedback may be useful to employ in future interventions.

Investigating how individuals responded to personalized health-related feedback was an important component of this study as it has been well documented that receiving tailored feedback can result in greater resonance and consequently result in desirable health behaviors [44-46]. Our study demonstrated that presenting feedback pertaining to an individual's glucose levels elicited significantly more brain activation within the middle frontal gyrus and subgyral compared with the behavioral feedback. These regions, anatomically positioned within Brodmann areas 9/10 and 47, respectively, have previously been associated with the actions of making personal moral judgments [47] and working memory [48], respectively. Previous studies have investigated messages promoting child vaccinations against measles-mumps-rubella and identified that highlighting the dangers of not vaccinating may actually be counterproductive [49]; therefore, findings are often highly dependent on the topic investigated. Future studies could investigate the role of self-affirmation, a construct suggested to increase individual sensitivity to health-risk information and incorporated in prior neuroimaging studies [29,41]. Self-affirmation essentially investigates how neural activity patterns vary to information after being exposed to personally important values (eg, friends, family, and religion). Given that the desirable outcome is for people to positively respond to health-related information, exposing a person to their personal values may provoke attention and enhance the importance of the information being given. Therefore, future investigation into whether self-affirmation could contribute to increasing the level of resonance toward personalized feedback and encourage positive behaviors may be crucial.

Behavior Change and Associations With Brain Activation

Our study identified a significant reduction of 29 minutes (or 47 minutes using inclinometry) in time spent sedentary from baseline to follow-up. Previous findings support this finding, having observed a 39.6 minute per day reduction in time spent sedentary [50]. However, no significant differences were observed for time spent in MVPA, light physical activity, or step count. Wearable technologies research to date has offered the suggestion that people can increase their activity levels having received feedback about behavior [51,52]. However, it must be acknowledged that physical activity, for example, has been categorized as a very complex behavior and no single metric can encapsulate a person's level of physical activity [53]. According to the literature, changes in behavior most likely occur when personalized health messages are presented in moments when action can be taken (eg, at midday to promote a walk following the consumption of lunch) [11]. Despite participants being presented personalized feedback, there are a multitude of reasons as to why they may or may not have changed their behavior during the follow-up period. Therefore, determining whether their behavior (change or no change) was because of the exposure to health-related information is truly unknown. However, emphasizing that the feedback was only briefly presented and within an unusual situation (ie, inside an MRI scanner) is warranted when comparing how people normally receive personalized feedback through wearable technologies. Further investigation could quantify or

contextualize the follow-up period to try and account for extraneous variables (eg, weather, holiday, illness) or consider the inclusion of a control group to provide more definitive findings.

In regard to the relationship of activation and subsequent behavior during the follow-up period, findings identified a positive partial correlation with minutes of MVPA. Previous studies have investigated behavior change subsequent to fMRI and have demonstrated positive associations between neural response (eg, to aversive smoking-related images) and smoking cessation [17,54]. For example, Falk and colleagues [41] identified that greater reductions in sedentary behavior aligned with greater activity in the ventromedial prefrontal cortex, suggesting that if people exhibited greater levels of activation in response to the visual stimuli, those individuals were subsequently more likely to be less sedentary. On a larger scale, identifying what stimuli (ie, health messages) evoke a positive prediction of behavior (eg, being less sedentary or more active) can inform the provision of effective public health messages. It could be suggested that, despite the observed association, being presented personalized feedback about health and behavior while inside an MRI scanner is not a normal environment. Consequently, alternate neuroimaging tools could be useful for future investigation within a free-living setting. For instance, individuals could obtain personalized feedback via a wearable device or a smartphone app while their neural activity is recorded by a portable electroencephalogram system via functional near infrared spectroscopy or by eye tracking (to monitor gaze patterns and fixations). Interestingly, eye tracking has previously been conducted on various health communication materials including both cigarette advertising [55] and nutrition labels [56].

Strengths and Limitations

Positioned at the intersection of a number of evolving interest areas, this interdisciplinary study offers a number of strengths. One strength was presenting the personalized feedback pertaining to both movement behaviors and physiology to participants. These components were objectively measured during baseline and follow-up using novel self-monitoring technologies, obtaining data to directly inform the feedback. In addition, the information that was presented in the fMRI tasks were designed based on feedback commonly presented via wearable devices or smartphone apps, reflecting what could be received in real-time in a real-world setting. Objective quantification of behavior at follow-up permitted the assessment of behavior following exposure and associations between neural activation and behavior.

Limitations of our study include the situation that participants viewed their glucose levels during baseline wear, an unavoidable situation given intentions to minimize data loss. This protocol confirms that participants had prior exposure to the glucose-related feedback subsequently presented during fMRI. However, to help try and account for this, analysis included the number of scans as a covariate because we thought the number of scans suggested the frequency with which participants viewed their glucose levels (eg, more scans equaled more exposure and so a greater awareness of their glucose levels). In addition, a lack of behavior change could be attributable to the sample that we recruited (ie, well educated and relatively healthy) and as such they could be profiled as a highly motivated audience who may not have viewed their behavior as in need of improvement. Furthermore, our unpowered sample size was another limitation, as we are unable to offer definitive interpretation of the findings. In addition, because of the number of people as active and inactive, we were unable to make any comparisons between groups of participants (eg, patterns of brain activation between those most active and least active). Finally, the pattern of neural activity observed and related psychological processes should be interpreted with caution due to the nature of reverse inference [57]. Future studies could investigate neural activity in polar groups of people classified by activity or time spent sedentary and repeat fMRI so patterns of brain activation are quantified before and after exposure to the feedback.

Conclusion

This multidisciplinary study highlighted that fMRI can be used to assess the neural response to personalized health feedback. In particular, greater activation in the prefrontal cortex during exposure to glucose compared with behavioral feedback was observed. A reduction in time spent sedentary and a negative association between the parameter estimates and subsequent minutes of MVPA were observed. Future research deploying behavioral feedback in parallel with physiological feedback to encourage positive behavior change is warranted.

Acknowledgments

The authors thank all the participants for taking part in this study. The authors would also like to thank Julie Thompson (senior radiographer, computed tomography/MRI, University Hospitals of Leicester) for her technical support and Dr Mark Orme (Centre for Exercise and Rehabilitation Science, Respiratory Biomedical Research Unit, Glenfield Hospital) and Dr Ruth Trethewey (School of Sport, Exercise and Health Sciences, Loughborough University) for their assistance with data collection. This work was supported by the School of Sport, Exercise and Health Sciences, the National Centre for Sport and Exercise Medicine (NCSEM) at Loughborough University, and the National Institute for Health Research (NIHR) Leicester Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, NCSEM England, the Department of Health, or the partners involved.

Conflicts of Interest

None declared.

References

1. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
2. Sedentary Behavior Research Network. Letter to the editor: standardized use of the terms "sedentary" and "sedentary behaviours". Appl Physiol Nutr Metab 2012 Jun;37(3):540-542. [doi: 10.1139/h2012-024] [Medline: 22540258]
3. Henson J, Yates T, Biddle SJH, Edwardson CL, Khunti K, Wilmot EG, et al. Associations of objectively measured sedentary behaviour and physical activity with markers of cardiometabolic health. Diabetologia 2013 May;56(5):1012-1020. [doi: 10.1007/s00125-013-2845-9] [Medline: 23456209]
4. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 1;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
5. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract 2014 Feb;103(2):137-149. [doi: 10.1016/j.diabres.2013.11.002] [Medline: 24630390]
6. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJH, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 04;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
7. Loveday A, Sherar LB, Sanders JP, Sanderson PW, Esliger DW. Technologies that assess the location of physical activity and sedentary behavior: a systematic review. J Med Internet Res 2015 Aug 05;17(8):e192 [FREE Full text] [doi: 10.2196/jmir.4761] [Medline: 26245157]
8. Bonander J, Gates S. Public health in an era of personal health records: opportunities for innovation and new partnerships. J Med Internet Res 2010 Aug 10;12(3):e33 [FREE Full text] [doi: 10.2196/jmir.1346] [Medline: 20699216]
9. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
10. Ledger D, McCaffrey D. Inside wearables: how the science of human behavior change offers the secret to long-term engagement.: Endeavor Partners URL: https://blog.endeavour.partners/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3?gi=9a190d5819b9 [accessed 2017-10-25] [WebCite Cache ID 6uUCtSRfS]
11. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
12. Fjeldsoe B, Neuhaus M, Winkler E, Eakin E. Systematic review of maintenance of behavior change following physical activity and dietary interventions. Health Psychol 2011 Jan;30(1):99-109. [doi: 10.1037/a0021974] [Medline: 21299298]
13. Lee VK, Harris LT. How social cognition can inform social decision making. Front Neurosci 2013 Dec 25;7:259 [FREE Full text] [doi: 10.3389/fnins.2013.00259] [Medline: 24399928]
14. Falk EB, Berkman ET, Mann T, Harrison B, Lieberman MD. Predicting persuasion-induced behavior change from the brain. J Neurosci 2010 Jun 23;30(25):8421-8424 [FREE Full text] [doi: 10.1523/JNEUROSCI.0063-10.2010] [Medline: 20573889]
15. Fishbein M, Cappella JN. The role of theory in developing effective health communications. J Communication 2006 Aug;56(s1):S1-S17. [doi: 10.1111/j.1460-2466.2006.00280.x]
16. Lieberman D. Social cognitive neuroscience. In: Lindzey G, editor. Handbook of Social Psychology. 3rd Edition. Mahwah: Lawrence Erlbaum; 2010.
17. Falk EB, Berkman ET, Whalen D, Lieberman MD. Neural activity during health messaging predicts reductions in smoking above and beyond self-report. Health Psychol 2011 Mar;30(2):177-185 [FREE Full text] [doi: 10.1037/a0022259] [Medline: 21261410]
18. Chua HF, Ho SS, Jasinska AJ, Polk TA, Welsh RC, Liberzon I, et al. Self-related neural response to tailored smoking-cessation messages predicts quitting. Nat Neurosci 2011 Apr;14(4):426-427 [FREE Full text] [doi: 10.1038/nn.2761] [Medline: 21358641]
19. Euston DR, Gruber AJ, McNaughton BL. The role of medial prefrontal cortex in memory and decision making. Neuron 2012 Dec 20;76(6):1057-1070 [FREE Full text] [doi: 10.1016/j.neuron.2012.12.002] [Medline: 23259943]
20. Wang Y. Anterior medial prefrontal cortex implements social priming of mimicry. Soc Cogn Affect Neurosci 2015 Apr;10(4):486-493 [FREE Full text] [doi: 10.1093/scan/nsu076] [Medline: 25009194]
21. Phan KL, Taylor SF, Welsh RC, Ho S, Britton JC, Liberzon I. Neural correlates of individual ratings of emotional salience: a trial-related fMRI study. Neuroimage 2004 Feb;21(2):768-780. [doi: 10.1016/j.neuroimage.2003.09.072] [Medline: 14980580]
22. Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF. Finding the self: an event-related fMRI study. J Cogn Neurosci 2002 Jul 01;14(5):785-794. [doi: 10.1162/08989290260138672] [Medline: 12167262]
23. Johnson SC, Baxter LC, Wilder LS, Pipe JG, Heiserman JE, Prigatano GP. Neural correlates of self-reflection. Brain 2002 Aug;125(Pt 8):1808-1814. [Medline: 12135971]

24. Chua HF, Liberzon I, Welsh RC, Strecher VJ. Neural correlates of message tailoring and self-relatedness in smoking cessation programming. Biol Psychiatry 2009 Jan 15;65(2):165-168 [FREE Full text] [doi: 10.1016/j.biopsych.2008.08.030] [Medline: 18926523]
25. Fishbein H, Triandis C, Kanfer F, Becker M. Factors influencing behavior and behavior change. In: Baum A, Revenson T, Singer J, editors. Handbook of Health Psychology. Mahwah: Lawrence Erlbaum; 2000:1-17.
26. Northoff G, Bermpohl F. Cortical midline structures and the self. Trends Cogn Sci 2004 Mar;8(3):102-107. [doi: 10.1016/j.tics.2004.01.004] [Medline: 15301749]
27. Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J. Self-referential processing in our brain—a meta-analysis of imaging studies on the self. Neuroimage 2006 May 15;31(1):440-457. [doi: 10.1016/j.neuroimage.2005.12.002] [Medline: 16466680]
28. Cascio N, Dal Cin S, Falk E. Health communications: predicting behavior change from the brain. In: Hall P, editor. Social Neuroscience and Public Health. New York: Springer; 2013:57-71.
29. Cooper N, Tompson S, O'Donnell M, Emily B. Brain activity in self-and value-related regions in response to online antismoking messages predicts behavior change. J Media Psychol 2015:1 [FREE Full text]
30. Dinh-Williams L, Mendrek A, Bourque J, Potvin S. Where there's smoke, there's fire: the brain reactivity of chronic smokers when exposed to the negative value of smoking. Prog Neuropsychopharmacol Biol Psychiatry 2014 Apr 03;50:66-73. [doi: 10.1016/j.pnpbp.2013.12.009] [Medline: 24361634]
31. Whelan ME, Morgan PS, Sherar LB, Orme MW, Esliger DW. Can functional magnetic resonance imaging studies help with the optimization of health messaging for lifestyle behavior change: a systematic review. Prev Med 2017 Jun;99:185-196. [doi: 10.1016/j.ypmed.2017.02.004] [Medline: 28214544]
32. Vecchiato G, Astolfi L, De Vico Fallani F, Toppi J, Aloise F, Bez F, et al. On the use of EEG or MEG brain imaging tools in neuromarketing research. Comput Intell Neurosci 2011;2011:3 [FREE Full text] [doi: 10.1155/2011/643489] [Medline: 21960996]
33. Wakefield MA, Durkin S, Spittal MJ, Siahpush M, Scollo M, Simpson JA, et al. Impact of tobacco control policies and mass media campaigns on monthly adult smoking prevalence. Am J Public Health 2008 Aug;98(8):1443-1450. [doi: 10.2105/AJPH.2007.128991] [Medline: 18556601]
34. Warburton E, Jamnik V, Bredin S, Gledhill N. The physical activity readiness questionnaire (PAR-Q) and electronic physical activity readiness medical examination (ePARmed-X). Heal Fit J Can 2011;4(2):3-23.
35. Aadland E, Ylvisåker E. Reliability of the Actigraph GT3X+ accelerometer in adults under free-living conditions. PLoS One 2015;10(8):e0134606 [FREE Full text] [doi: 10.1371/journal.pone.0134606] [Medline: 26274586]
36. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
37. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. 24 hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2015 Oct 17;48(3):457-465. [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
38. Bailey T, Bode BW, Christiansen MP, Klaff LJ, Alva S. The performance and usability of a factory-calibrated flash glucose monitoring system. Diabetes Technol Ther 2015 Nov;17(11):787-794 [FREE Full text] [doi: 10.1089/dia.2014.0378] [Medline: 26171659]
39. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002 Jan;15(1):273-289. [doi: 10.1006/nimg.2001.0978] [Medline: 11771995]
40. Enax L, Hu Y, Trautner P, Weber B. Nutrition labels influence value computation of food products in the ventromedial prefrontal cortex. Obesity (Silver Spring) 2015 Apr;23(4):786-792 [FREE Full text] [doi: 10.1002/oby.21027] [Medline: 25755174]
41. Falk EB, O'Donnell MB, Cascio CN, Tinney F, Kang Y, Lieberman MD, et al. Self-affirmation alters the brain's response to health messages and subsequent behavior change. Proc Natl Acad Sci U S A 2015 Feb 17;112(7):1977-1982 [FREE Full text] [doi: 10.1073/pnas.1500247112] [Medline: 25646442]
42. Newman-Norlund RD, Thrasher JF, Fridriksson J, Brixius W, Froeliger B, Hammond D, et al. Neural biomarkers for assessing different types of imagery in pictorial health warning labels for cigarette packaging: a cross-sectional study. BMJ Open 2014 Dec 31;4(12):e006411 [FREE Full text] [doi: 10.1136/bmjopen-2014-006411] [Medline: 25552613]
43. Tognoli E, Kelso JAS. The coordination dynamics of social neuromarkers. Front Hum Neurosci 2015;9:563 [FREE Full text] [doi: 10.3389/fnhum.2015.00563] [Medline: 26557067]
44. Strecher VJ, Kreuter M, Den Boer DJ, Kobrin S, Hospers HJ, Skinner CS. The effects of computer-tailored smoking cessation messages in family practice settings. J Fam Pract 1994 Sep;39(3):262-270. [Medline: 8077905]
45. Skinner CS, Strecher VJ, Hospers H. Physicians' recommendations for mammography: do tailored messages make a difference? Am J Public Health 1994 Jan;84(1):43-49. [Medline: 8279610]
46. Brug J, Steenhuis I, van Assema P, de Vries H. The impact of a computer-tailored nutrition intervention. Prev Med 1996;25(3):236-242. [doi: 10.1006/pmed.1996.0052] [Medline: 8781000]

47. Greene J, Haidt J. How (and where) does moral judgment work? Trends Cogn Sci 2002 Dec 01;6(12):517-523. [Medline: 12475712]
48. D'Esposito M, Aguirre GK, Zarahn E, Ballard D, Shin RK, Lease J. Functional MRI studies of spatial and nonspatial working memory. Brain Res Cogn Brain Res 1998 Jul;7(1):1-13. [Medline: 9714705]
49. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: a randomized trial. Pediatrics 2014 Apr;133(4):e835-e842 [FREE Full text] [doi: 10.1542/peds.2013-2365] [Medline: 24590751]
50. Chu AHY, Ng SHX, Tan CS, Win AM, Koh D, Müller-Riemenschneider F. A systematic review and meta-analysis of workplace intervention strategies to reduce sedentary time in white-collar workers. Obes Rev 2016 May;17(5):467-481. [doi: 10.1111/obr.12388] [Medline: 26990220]
51. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
52. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
53. Thompson D, Peacock O, Western M, Batterham AM. Multidimensional physical activity: an opportunity, not a problem. Exerc Sport Sci Rev 2015 Apr;43(2):67-74 [FREE Full text] [doi: 10.1249/JES.0000000000000039] [Medline: 25607280]
54. Jasinska AJ, Chua HF, Ho SS, Polk TA, Rozek LS, Strecher VJ. Amygdala response to smoking-cessation messages mediates the effects of serotonin transporter gene variation on quitting. Neuroimage 2012 Mar;60(1):766-773 [FREE Full text] [doi: 10.1016/j.neuroimage.2011.12.064] [Medline: 22245645]
55. Krugman M, Fox R, Fletcher J, Fischer P, Rojas T. Do adolescents attend to warnings in cigarette advertising: an eye-tracking approach. J Advert Res 1994:39.
56. Oliveira D. Consumers' attention to functional food labels: insights from eye-tracking and change detection in a case study with probiotic milk. Food Sci Technol 2016;68:160-167 [FREE Full text]
57. Poldrack RA. Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 2006 Feb;10(2):59-63. [doi: 10.1016/j.tics.2005.12.004] [Medline: 16406760]

Abbreviations

ANCOVA: analysis of covariance
cpm: counts per minute
fMRI: functional magnetic resonance imaging
FSPGR: fast spoiled gradient echo
FEW: family-wise error
GLM: general linear model
HbA1c: hemoglobin A1c
hem: hemisphere
LVPA: light-to-vigorous physical activity
MET: metabolic equivalent of task
MNI: Montreal Neurological Institute
mPFC: medial prefrontal cortex
MVPA: moderate-to-vigorous physical activity
NCSEM: National Centre for Sports and Exercise Medicine
NIHR: National Institute for Health Research
ROI: region of interest
SPM: statistical parametric mapping
WFU: Wake Forest University

Edited by G Eysenbach; submitted 01.09.17; peer-reviewed by A Cross, N Gell; comments to author 27.09.17; revised version received 06.10.17; accepted 09.10.17; published 08.11.17.

Please cite as:
Whelan ME, Morgan PS, Sherar LB, Kingsnorth AP, Magistro D, Esliger DW
Brain Activation in Response to Personalized Behavioral and Physiological Feedback From Self-Monitoring Technology: Pilot Study
J Med Internet Res 2017;19(11):e384
URL: http://www.jmir.org/2017/11/e384/
doi:10.2196/jmir.8890
PMID:29117928

Original Paper

Tailored Web-Based Interventions for Pain: Systematic Review and Meta-Analysis

Geraldine Martorella[1,2,3*], RN, PhD; Madalina Boitor[4*], RN; Melanie Berube[4,5,6*], MSc; Suzanne Fredericks[7], RN, PhD; Sylvie Le May[3,8,9], RN, PhD; Céline Gélinas[3,4,10,11,12*], RN, PhD

[1]College of Nursing, Florida State University, Tallahassee, FL, United States
[2]Tallahassee Memorial Hospital Center for Research and Evidence-Based Practice, Tallahassee, FL, United States
[3]Quebec Nursing Intervention Research Network (RRISIQ), Montreal, QC, Canada
[4]Ingram School of Nursing, McGill University, Montreal, QC, Canada
[5]Department of Trauma, Centre Integré Universitaire du Nord de l'Île de Montréal, Hôpital du Sacré-Cœur de Montréal, Montreal, QC, Canada
[6]Department of Nursing, Centre Integré Universitaire du Nord de l'Île de Montréal, Hôpital du Sacré-Cœur de Montréal, Montreal, QC, Canada
[7]Daphne Cockwell School of Nursing, Ryerson University, Toronto, ON, Canada
[8]Faculté des sciences infirmières, Université de Montréal, Montreal, QC, Canada
[9]Centre de recherche, Centre hospitalier universitaire Ste Justine, Montreal, QC, Canada
[10]Center for Nursing Research, Jewish General Hospital, McGill University, Montreal, QC, Canada
[11]Alan Edwards Centre for Research on Pain, McGill University, Montreal, QC, Canada
[12]Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, QC, Canada
[*]these authors contributed equally

Corresponding Author:
Geraldine Martorella, RN, PhD
College of Nursing
Florida State University
Vivian M Duxbury Hall
98 Varsity Way
Tallahassee, FL, 32306
United States
Phone: 1 850 644 6028
Email: gmartorella@fsu.edu

Abstract

Background: Efforts have multiplied in the past decade to underline the importance of pain management. For both acute and chronic pain management, various barriers generate considerable treatment accessibility issues, thereby providing an opportunity for alternative intervention formats to be implemented. Several systematic reviews on Web-based interventions with a large emphasis on chronic pain and cognitive behavioral therapy have been recently conducted to explore the influence of these interventions on pain management However, to our knowledge, the specific contribution of tailored Web-based interventions for pain management has not been described and their effect on pain has not been evaluated.

Objective: The primary aim of this systematic review was to answer the following research question: What is the effect of tailored Web-based pain management interventions for adults on pain intensity compared with usual care, face-to-face interventions, and standardized Web-based interventions? A secondary aim was to examine the effects of these interventions on physical and psychological functions.

Methods: We conducted a systematic review of articles published from January 2000 to December 2015. We used the DerSimonian-Laird random effects models with 95% confidence intervals to calculate effect estimates for all analyses. We calculated standardized mean differences from extracted means and standard deviations, as outcome variables were measured on different continuous scales. We evaluated 5 different outcomes: pain intensity (primary outcome), pain-related disability, anxiety, depression, and pain catastrophizing. We assessed effects according to 3 time intervals: short term (<1 month), medium term (1-6 months), and long term (6-12 months).

Results: After full-text review, we excluded 31 articles, resulting in 17 eligible studies. Only 1 study concerned acute pain and was removed from the meta-analysis, resulting in 16 studies available for quantitative assessment. Compared with standard care

or a waiting list, tailored Web-based intervention showed benefits immediately after, with small effect sizes (<0.40) for pain intensity (10 randomized controlled trials [RCTs], n=1310, *P*=.003) and pain-related disability (6 RCTs, n=953, *P*<.001). No other improvements were observed at follow-up in the medium and long terms. Compared with the active control group, no improvements were found for the primary outcome (pain intensity) or any of the outcomes except for a small effect size on pain catastrophizing (2 RCTs, n=333, *P*<.001) immediately after the intervention.

Conclusions: Tailored Web-based interventions did not prove to be more efficacious than standardized Web-based interventions in terms of pain intensity, pain-related disability, anxiety, and depression. An interesting finding was that some efficacy was shown on pain catastrophizing compared with active control interventions. Considering the diversity of approaches used in tailored Web-based interventions for chronic pain management, their efficacy is yet to be explored. Moreover, their contribution to acute pain management is embryonic.

Trial Registration: International prospective register of systematic reviews (PROSPERO): CRD42015027669; http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42015027669 (Archived by WebCite at http://www.webcitation.org/6uneWAuyR)

(J Med Internet Res 2017;19(11):e385) doi:10.2196/jmir.8826

KEYWORDS

Web-based intervention; tailored intervention; pain management; chronic pain; acute pain; review; systematic review; meta-analysis

Introduction

Efforts have been made in the past decade to underscore the importance of pain management and its status as a human right [1-3]. Acute pain is a widespread issue. Annually, 300 million surgeries are performed worldwide, with a third occurring in the United States, and have resulted in approximately 80% of patients reporting pain [4]. As well, 70% of emergency departments visits are related to acute pain [4]. Furthermore, it is estimated that approximately 100 million adults in the United States have chronic pain [3], with 25.3 million adults experiencing daily pain [5]. Although caseloads and wait times are difficult to estimate, particularly in the United States, it is generally recognized that treatment availability for chronic pain patients is scarce [6,7]. Significant barriers such as time, cost, and distance generate considerable treatment accessibility issues [3] and inhibit the improvement of pain management, thereby providing an opportunity for alternative formats to face-to-face interventions to be implemented [8-10].

Over the past decade, more Web-based interventions for pain management have been developed and, as opposed to non-Web-based interventions, they have been shown to positively influence health behaviors [11,12]. Many terms have been used interchangeably to qualify Web-based interventions that facilitate the implementation of self-management health-related interventions. Here the term Web-based interventions refers to Barak et al's definition [13]:

> *...a primarily self-guided intervention program that is executed by means of a prescriptive online program operated through a website and used by consumers seeking health- and mental-health[-]related assistance. The intervention program itself attempts to create positive change and or improve/enhance knowledge, awareness, and understanding via the provision of sound health-related material and use of interactive Web-based components.*

This definition is composed of three types of health-related interventions: educational, self-guided therapeutic, and human-supported therapeutic interventions [13].

Several systematic reviews on Web-based interventions for pain, with a large emphasis on chronic pain issues and cognitive behavioral therapy (CBT) interventions, have been conducted to explore their influence on pain management [14-19]. Overall, their authors concluded that results were promising in terms of pain reduction, and improvement of functional and emotional well-being. They also underlined that it is still unknown as to which type of patients, according to sociodemographic and clinical characteristics, would benefit most from a Web-based approach [14,15]. A small positive effect was found on pain reduction, but results remain inconclusive considering the small sample sizes for several studies, high dropout rates, and heterogeneity related to assessment tools used and times of measurement selected, type of pain-related diseases, and interventions (eg, content, format, dose), but also the lack of diversity in patients (eg, mainly women, white, and college educated) [10,14,15,17,19]. Nonetheless, none of the systematic reviews looking at Web-based interventions for pain management specifically addressed the contribution of tailoring ingredients.

Experts in health behavior change have shown that conveying health information without considering individual differences may inhibit behavior change [20-26]. Tailoring strategies respond to this concern and computed algorithms can facilitate the implementation of this approach in terms of both accessibility and level of refinement. Tailoring is defined as a process for creating individualized communications using personal data related to health outcomes in order to meet individual needs [21,23,25,27,28]. Three mechanisms have been highlighted [21,29]: (1) personalization, which helps increase the perceived meaningfulness of the message by creating the impression that the message was designed specifically for the individual [21]; (2) feedback, which directs the attention of the individual to their own characteristics or behaviors that they need to address, improve, or change [21]; and (3) adaptation or content matching, which refers to creating content packages

that are pertinent to an individual and are selected based on known determinants of the targeted behavior [21]. Clinically relevant results, such as adopting a healthy lifestyle or adhering to medication, and statistically significant effect sizes of tailored Web-based interventions have been recognized for health behavior change among diverse populations facing chronic disease [20,29,30]. However, the contribution of tailored Web-based interventions for pain management has not been described, and their specific effect on pain has not been evaluated.

Therefore, this systematic review aimed to answer the following research question: What is the effect of tailored Web-based pain management interventions for adults on pain intensity compared with usual care, face-to-face interventions, and standardized Web-based interventions? We also examined secondary outcomes related to the effects of these interventions on physical and psychological functions.

Methods

This systematic review protocol has been developed based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for reporting systematic reviews evaluating health care interventions [31-33]. The detailed protocol was published [34] and registered with the international prospective register of systematic reviews (no CRD42015027669).

Inclusion Criteria

We include solely randomized controlled trials (RCTs) in this systematic review. Other inclusion criteria were studies with patients 18 years of age or older and experiencing any type of pain (acute, ie, <3 months; chronic, ie, 3-6 months and beyond [35]). We selected studies involving Web-based interventions for pain management including at least one of the three tailoring strategies (personalization, feedback, or adaptation) [21,29]. Eligible comparators were (1) a passive control group (ie, participants receiving usual medical and nursing care or being on a pain clinic waitlist), and (2) active control group (eg, face-to-face educational or psychological intervention or Web-based standardized intervention) [14,15].

We selected outcomes according to pain clinical trials recommendations [36-38]. Pain intensity was a mandatory outcome for the study to be included in this systematic review. The timeline of outcomes included measures before and immediately after treatment and at follow-up. To reduce selection bias, if articles were published in languages other than English or French, we reviewed the English abstract to determine whether the study should be translated and included. This was the case for an article published in German identified as eligible [39] and translated by a member of the team (MaB).

Search Strategy

We searched MEDLINE, Embase, CINAHL, PsycINFO, Web of Science, and the Cochrane Library for articles published from January 2000 to December 2015. Reviewing the reference lists of relevant articles and previous systematic reviews helped identify 9 additional articles. Moreover, an experienced research librarian used subject headings to avoid missing nonindexed concepts. Search terms were “pain,” “pain management,” “program,” “intervention,” “Internet,” “Internet-based,” “online,” “Web-based,” and “mobile OR mobile applications” [34].

Screening and Selection of Studies

Eligibility was assessed independently in an unblinded standardized manner by 2 team members (GM, CG) and results were then compared. Titles and abstracts were screened. If a trial was potentially eligible, the full text was reviewed. The 2 reviewers are researchers in the field of pain with a clinical background in nursing. Disagreements between reviewers at the full-text level were discussed until consensus was reached.

Data Extraction and Management

Data were extracted independently by 2 teams of 2 reviewers (GM and MaB; CG and MeB) composed of 1 doctoral student and 1 researcher using the software DistillerSR v2 (Evidence Partners Inc). We developed the data extraction form we used based on the Cochrane Consumers and Communication Group’s data extraction template [40], pilot tested it with 5 articles and refined it accordingly (eg, number of comparator arms, time points for postintervention assessments). Disagreements between reviewers were resolved by discussion between the 2 teams until 100% agreement was reached. We requested missing data such as means and standard deviations regarding the outcome variables from authors. Extracted data included sample size, sample demographics, dropout rate, number and type of study groups, type and location of pain, inclusion and exclusion criteria, study setting, type of Web-based intervention (ie, setting, mode, dose, contact with therapist, hybrid format), tailoring strategy (ie, personalization, feedback, adaptation), comparator (ie, passive control group vs active control group), type of pain intensity measure, pain-related disability and psychological well-being outcomes, and times of measurement.

Data Assessment and Synthesis

Risk of bias for the 17 selected studies was assessed by the 2 teams of reviewers using the Cochrane Collaboration tool [41-43], and any discrepancies between reviewers were discussed between the 2 teams. The report of the risk-of-bias assessment is presented in the Results section.

We used Review Manager (RevMan 5.3; Cochrane Collaboration) software [44] for statistical analysis. We used the DerSimonian-Laird random effects models with 95% confidence intervals to calculate effect estimates for all analyses. We calculated standardized mean differences (SMDs) from extracted means and standard deviations, as outcome variables were measured on different continuous scales. We planned subgroup analyses by the comparator (ie, passive or active control) and type of pain (ie, acute or chronic). The primary outcome was pain intensity measured using a self-report measure such as the numeric rating scale (NRS) (eg,) or visual analog scale (VAS) (eg, 0-10 cm or 0-100 mm). We converted scores reported on a 0-100 mm VAS to 0-10 NRS scores for the purpose of data analysis [45]. If authors provided data for least, average, and worst pain intensity as measured by the Brief Pain Inventory, we used average scores for data analysis. Secondary outcomes, if available, were pain-related disability (eg, Brief

Pain Inventory; Roland Morris Disability Questionnaire), and psychological well-being (eg, Hospital Anxiety and Depression Scale; Pain Catastrophizing Scale). In an effort to decrease the heterogeneity found in pain-related disability and psychological well-being measures, we included tools measuring the same construct to calculate SMDs. Given the variability in follow-up assessments across the included studies, we report outcomes according to 3 different time intervals: (1) short-term effect: immediately after or within a month after intervention, (2) medium-term effect: up to 6 months after completion of the intervention, (3) long-term effects: over 6 months after completion of the intervention. We evaluated between-study variability using the method proposed by Higgins et al [42]. We considered an I^2 statistic above 50% to indicate high heterogeneity, values between 25% and 50% to indicate moderate heterogeneity, and those below 25% to indicate low heterogeneity.

Results

After a full-text review of 48 articles, we excluded 31 for the following reasons: 3 studies were not RCTs, 18 studies did not involve 1 of the tailoring mechanisms, and 10 studies did not measure pain intensity using a VAS or NRS or a pain index calculated with scores obtained on an NRS over a period of time. We included 17 studies in the qualitative synthesis and 16 in the meta-analysis (see Figure 1). Of note, the effects of 1 intervention were described in 2 articles: short- and medium-term effects [46] and long-term effects [47].

Study Characteristics

As Table 1 presents, we included 17 studies, 16 of which were performed in the chronic pain context [39,46-61] and 1 in the acute pain context in the postcardiac surgery phase [62]. Concerning studies conducted in the chronic pain context (n=16), 6 included individuals with back pain [39,50-52,55,59], 3 included individuals with other specific pain sites (head as well as hips and knees) [48,49,61], 5 included individuals with multiple pain sites or widespread pain [46,47,53,54,58,60], and 2 included individuals with chronic disease(s) (ie, heart disease, lung disease, type 2 diabetes, mobility difficulty, chronic musculoskeletal pain, and depression) [56,57]. Considering that only 1 study was performed in the acute care context, we included only data from chronic pain studies in the meta-analysis.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart. NRS: numeric rating scale; RCT: randomized controlled trial; VAS: visual analog scale.

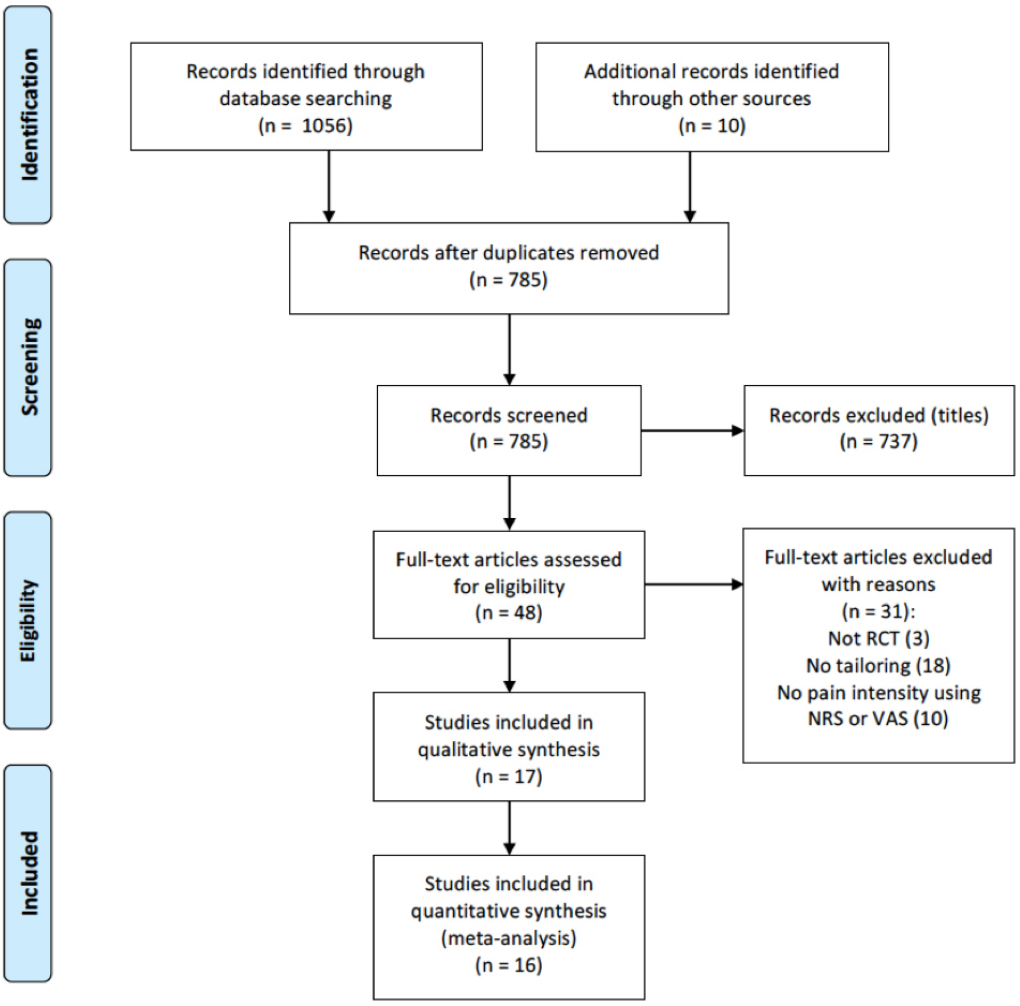


Table 1. Characteristics of the 17 eligible studies.

First author, year, reference	Country	Sample size (n)	Lost to follow-up[a] (%)	Age in years, mean (SD)	Female (%)	Type of pain	Pain location
Andersson, 2003 [48]	Sweden	44	45.5	40.3 (NR[b])	81.7	Chronic	Headache
Bossen, 2013 [49]	Netherlands	199	15.6	62 (5.7)	64.8	Chronic	Hips, knees
Buhrman, 2004 [50]	Sweden	56	8.9	44.6 (10.4)	62.5	Chronic	Back
Carpenter, 2012 [51]	United States	141	7.1	42.5 (10.3)	83	Chronic	Low back
Chiauzzi, 2010 [52]	United States	209	11	46.1 (11.9)	67	Chronic	Back
Dear, 2013 [53]	Australia	63	4.8	49 (13)	86	Chronic	Multiple
Dear, 2015 [54]	Australia	490	14.1	50 (13)	80	Chronic	Multiple
Krein, 2013 [55]	United States	229	9.6	51.5 (NR)	12.5	Chronic	Back
Kristjánsdóttir, 2013 [46,47]	Norway	140	40	44.2 (NR)	100	Chronic	General
Leveille, 2009 [56]	United States	241	22.8	52 (12)	57	Chronic	Chronic disease related
Lorig, 2006 [57]	United States	958	18.2	57.5 (10.9)	71	Chronic	Chronic disease related
Lorig, 2008 [58]	United States	855	25	52 (11.6)	90	Chronic	Arthritis/ fibromyalgia
Martorella, 2012 [62]	Canada	60	13.3	64.6 (8.2)	21	Acute	Surgical site
Moessner, 2012 [59]	Germany	75	44	45.9 (NR)	57	Chronic	Back
Moessner, 2014 [39]	Germany	334	21.3	47.36 (9.89)	63.8	Chronic	Back
Shigaki, 2013 [60]	United States	108	13.9	49.8 (NR)	92.3	Chronic	General
Ström, 2000 [61]	Sweden	102	56	36.7 (NR)	68	Chronic	Head

[a]Rates are calculated based on the number of randomly assigned participants who completed a posttreatment questionnaire (time points may vary within studies).

[b]NR: not reported.

The total number of participants entering chronic pain trials was 4103 (mean 256.4 participants per study, SD 270.8, median 170, interquartile range 81.8-310.8). All studies described the total number of participants providing data at the end of the interventions. The mean completion rate for studies that provided such data was 77.6%, with the proportion of completers ranging across studies from 44% to 95.2%. The mean age of participants entering the studies was 48.2 years (SD 6.3, range 36.7-62.0, median 48.2, interquartile range 44.3-51.9). The average proportion of female participants was 71%.

A total of 5 studies used usual care as the comparator arm [39,55,57-59], 6 used a waiting list [49-51,53,60,61], and 5 included active controls, in which participants received Web-based information, psychological support, or standardized CBT without in-person contact with a therapist [46-48,52,54,56]. All studies used 1 comparator arm, except for 1 study that used 3: standardized CBT and tailored psychological support with optional health professional contact; standardized CBT and tailored psychological support without health professional contact; and a waiting list [54]. The interventions included 5 to 18 sessions over a period of 3 weeks to 12 months, and most of them were provided weekly.

A total of 13 studies evaluated Web-based interventions using a CBT or behavioral approach [39,46-54,57-60], which was combined with an additional approach (ie, education, relaxation, mindfulness therapy, and motivational and psychological support) in 9 studies [46-48,50-54,57,58]. Exercise, motivation, coaching, education, and relaxation approaches were used in the 3 trials that did not use CBT [55,56,61]. A total of 7 interventions were delivered with a hybrid mode of delivery precisely combining Internet and in-person contact with a health professional over the telephone or face-to-face [39,46-48,50,53,54,60]. Feedback (mediated or not) was used in every intervention. Content matching was used in half of the studies [39,46,47,49,52,55,57-59]. Personalization was difficult to assess given the lack of a detailed description of interventions and platforms in research articles, but also because it was embedded in the 2 other tailoring mechanisms. Table 2 summarizes the approaches and dosage of tailored Web-based interventions and their comparator.

Table 2. Description of tailored Web-based interventions and their comparator.

First author, year, reference	Web-based tailored intervention				Comparator	
	Approach	Format	Duration	Frequency	Approach	Type
Andersson, 2003 [48]	Relaxation + CBT[a]	Hybrid: telephone	6 weeks	1/week (6 sessions)	Same as intervention	Active (Web based)
Bossen, 2013 [49]	Behavioral graded activity	Hybrid: N/A[b]	9 weeks	1/week (9 sessions)	N/A	Waiting list
Buhrman, 2004 [50]	CBT + relaxation, exercise, and stretching	Hybrid: telephone	6 weeks	1/week (6 sessions)	N/A	Waiting list
Carpenter, 2012 [51]	CBT + relaxation, mindfulness	Hybrid: N/A	3 weeks	2/week (6 sessions)	N/A	Waiting list
Chiauzzi, 2010 [52]	CBT + motivational + educational (wellness, lifestyle)	Hybrid: N/A	4 weeks + 5 monthly boosters (6 months)	2/week (8 sessions)	Emailed back pain information guide	Active
Dear, 2013 [53]	CBT + educational (sleep hygiene)	Hybrid: telephone	8 weeks	Every 7-10 days (5 sessions)	N/A	Waiting list
Dear, 2015 [54]	CBT + psychological approach	Hybrid: telephone	8 weeks	Every 7-10 days (5 sessions)	Same as intervention	Active (Web based) and waitlist
Krein, 2013 [55]	Exercise, motivational	Hybrid: N/A	12 months	Weekly feedback, reminder (daily sessions)	Wearing a pedometer and reminders to upload data	Usual care
Kristjánsdóttir, 2013 [46,47]	CBT + ACT[c] + mindfulness	Hybrid: face-to-face	4 weeks	5/week (20 sessions)	Information website	Active
Leveille, 2009 [56]	Coaching + educational (disease specific)	Hybrid: N/A	4 weeks	N/A	URL links provided to patients: home pages for the US Department of Health and Human Services and the Centers for Disease Control and Prevention	Active
Lorig, 2006 [57]	CBT + educational (nutrition, medication)	Hybrid: N/A	6 weeks	3/week (18 sessions)	N/A	Usual care
Lorig, 2008 [58]	CBT + educational (nutrition, medication)	Hybrid: N/A	6 weeks	3/week (18 sessions)	N/A	Usual care
Martorella, 2012 [62]	CBT	Hybrid: face-to-face	30 minutes + 2 boosters (5-10 minutes)	1 session before surgery	N/A	Usual care
Moessner, 2012 [59]	Behavioral	Hybrid: N/A	12-15 weeks	1/week (12-15 sessions)	N/A	Usual care
Moessner, 2012 [39]	CBT	Hybrid: face-to-face	12-15 weeks	1/week (12-15 sessions)	N/A	Usual care
Shigaki, 2013 [60]	CBT	Hybrid: telephone	10 weeks	1/week (10 sessions)	N/A	Waiting list
Ström, 2000 [61]	Relaxation	Hybrid: N/A	6 weeks	1/week (6 sessions)	N/A	Waiting list

[a]CBT: cognitive behavioral therapy.

[b]N/A: not applicable.

[c]ACT: acceptance and commitment therapy.

Table 3. Risk of bias within studies according to reviewers.

First author, year, reference	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias
Andersson, 2003 [48]	Unclear	Unclear	Low	Low	High	Unclear	Low
Bossen, 2013 [49]	Low	Low	High	High	Low	Low	Low
Buhrman, 2004 [50]	Low	Unclear	High	High	Low	Low	Low
Carpenter, 2012 [51]	Low	Unclear	High	High	High	Low	Low
Chiauzzi, 2010 [52]	Low	Unclear	Unclear	Unclear	Low	Low	Low
Dear, 2013 [53]	Unclear	Unclear	High	High	Low	Low	Low
Dear, 2015 [54]	Low	Low	High	High	Low	Low	Low
Krein, 2013 [55]	Low	Low	High	High	Low	Low	Low
Kristjánsdóttir, 2013 [46,47]	Low	Low	High	High	Low	Low	Low
Leveille, 2009 [56]	Unclear	Low	High	High	High	Low	Low
Lorig, 2006 [57]	Unclear	Unclear	High	High	High	Low	Low
Lorig, 2008 [58]	Unclear	Low	High	High	High	Low	Low
Martorella, 2012 [62]	Low	Low	High	High	Low	Low	Low
Moessner, 2012 [59]	Unclear	Unclear	High	High	Low	Unclear	Unclear
Moessner, 2014 [39]	Unclear	Unclear	High	High	Low	Unclear	Unclear
Shigaki, 2013 [60]	Unclear	Unclear	High	High	High	Low	Unclear
Ström, 2000 [61]	Unclear	Unclear	High	High	High	Unclear	Low

Risk-of-Bias Assessment

A total of 6 potential biases were evaluated according to the Cochrane Collaboration tool: selection, performance, detection, attrition, reporting, and other [42]. Regarding selection bias, randomized sequence generation presented an unclear risk for about half of the studies (n=9) [39,48,53,56-61] and a low risk of bias for the other half (n=8). Unclear risk was mainly related to insufficient information on the sequence generation. It was also the case for allocation concealment, which we judged to be unclear for 10 studies [39,48,50-53,57,59-61]. In terms of performance bias, all studies except for 2 (low and unclear risks) [48,52] presented a high risk in regard to blinding of participants, which was also the case for detection bias. These 2 biases are almost inevitable with this type of intervention, especially compared with a waitlist, as the group assignment is easy to guess for participants who are the actual outcome assessors. Indeed, most of the time, outcomes were measured through self-administered online questionnaires. The study by Andersson et al [48] presented a low risk because both groups received the same Web-based intervention and completed online questionnaires, with the only difference being a telephone contact added in the experimental group. Another study [51] used this strategy and provided the same intervention to both groups but not at the same time. The control group had to complete questionnaires before receiving the intervention, which ended up presenting a risk of bias. The risk in the study by Chiauzzi et al [52] remained unclear because of a lack of information. Blinding could have occurred even though the control group did not access a website but received a guide to consult. We found attrition risk of bias in 7 studies that presented a high risk regarding incomplete data [48,51,56-58,60,61]. The reasons for attrition were sometimes not explained, and information was lacking regarding the method for handling missing data. We judged selective reporting as low risk except in 4 studies (unclear) [39,59]. A total of 9 studies had registered protocols [46,47,49,53-56,58,60,62] and, for those without, we judged a low risk for selective reporting given the consistency between methods and results. Other risks of bias were low, except in 3 studies (unclear), in which some information was lacking regarding the methods and attrition. Table 3 summarizes the risk-of-bias assessment for each study. We constructed funnel plots for only 1 analysis due to the small number of studies included in each analysis (n<10) [63,64].

Effects of Interventions

We included 16 studies in the quantitative analysis, having removed the single study on acute pain [62]. We conducted 2 main meta-analyses based on the type of comparator: tailored Web-based intervention versus standard care or waitlist control, and tailored Web-based intervention versus active control. For both meta-analyses, when possible, we analyzed outcomes at 3 different time points (ie, short, medium, and long term). Table 4 reports the tools used for outcome assessment and timeline per study. We included the study with 3 control groups (ie, optional contact, no contact with therapist, and standard care or waitlist control) [54] in both meta-analyses, such that the control group that received the standardized Web-based intervention without contact was included in the tailored Web-based intervention versus active control, and the waitlist control group was included in the tailored Web-based intervention versus standard care or waitlist control.

Table 4. Assessment tools and timing.

First author, year, reference	Pain (with score range)	Pain-related disability	Psychological well-being	Timing of posttreatment assessment		
				Short	Medium	Long
Andersson, 2003 [48]	0-5 NRS[a,b]	HDI[c]	HADS[d]	X		
Bossen, 2013 [49]	0-10 NRS	PASE[e], KOOS/HOOS[f]	HADS	X		X
Buhrman, 2004 [50]	0-100 VAS[g]	MPI[h]	HADS	X	X	
Carpenter, 2012 [51]	0-10 NRS	RMDQ[i]	PCS[j]	X		
Chiauzzi, 2010 [52]	0-10 NRS	BPI[k]	PCS, DASS[l]	X	X	
Dear, 2013 [53]	0-10 NRS	RMDQ	PHQ-9[m] (depression), GAD-7[n] (anxiety)	X	X	
Dear, 2015 [54]	0-10 NRS	RMDQ	PHQ-9 (depression), GAD-7 (anxiety)	X	X	
Krein, 2013 [55]	0-10 NRS	RMDQ	PCS	X		
Kristjánsdóttir, 2013 [46,47]	0-100 VAS	FIQ[o]	PCS	X	X	X
Leveille, 2009 [56]	0-10 NRS	N/A[p]	N/A		X	
Lorig, 2006 [57]	0-10 VAS	IIS[q], HAI[r]	N/A		X	X
Lorig, 2008 [58]	0-10 VAS	ALS[s], HAQ[t]	N/A		X	X
Martorella, 2012 [62]	0-10 NRS	BPI	PCS	X		
Moessner, 2012 [59]	0-10 NRS	RMDQ	HADS	X	X	
Moessner, 2014 [39]	0-10 NRS	RMDQ	N/A	X	X	X
Shigaki, 2013 [60]	0-100 NRS with 5-point increment	AIMS[u]	CES-D[v] (depression)	X		X
Ström, 2000 [61]	0-100 NRS	HDI	BDI[w] (depression)	X		

[a]Headache index: means of noted pain intensity for each day summed, divided by the total number of registration days.
[b]NRS: numeric rating scale.
[c]HDI: Headache Disability Inventory.
[d]HADS: Hospital Anxiety and Depression Scale.
[e]PASE: Physical Activity Scale for the Elderly.
[f]KOOS/HOOS: Knee Osteoarthritis Outcome Score/Hip Injury Osteoarthritis Outcome Score.
[g]VAS: visual analog scale.
[h]MPI: Multidimensional Pain Inventory.
[i]RMDQ: Roland Morris Disability Questionnaire.
[j]PCS: Pain Catastrophizing Scale.
[k]BPI: Brief Pain Inventory.
[l]DASS: Depression Anxiety Stress Scale.
[m]PHQ-9: Patient Health Questionnaire.
[n]GAD-7: Generalized Anxiety Disorder.
[o]FIQ: Fibromyalgia Impact Questionnaire.
[p]N/A: not applicable.
[q]IIS: Illness Intrusiveness Rating Scale.
[r]HAI: Health Assessment Instrument.
[s]ALS: Activities Limitation Scale.
[t]HAQ: Health Assessment Questionnaire.
[u]AIMS: Arthritis Impact Measurement Scale.
[v]CES-D: Center for Epidemiologic Studies Depression Scale.
[w]BDI: Beck Depression Inventory.

We removed the control group that received the standardized Web-based intervention with optional contact from the meta-analysis, as it was impossible to evaluate the amount of contact with a therapist that was received. Overall, we included 16 studies in the 2 main meta-analyses: 11 in the tailored Web-based intervention versus standard care or waitlist control [39 49-51,53,55,57-61], and 4 in the tailored Web-based intervention versus active control [46-48,52,56]; we included 1 in both [54].

Effects of Tailored Web-Based Interventions Versus Standard Care

Pain

We entered 10 studies (n=1310) into an analysis of the short-term effect of Web-based tailored interventions on pain intensity. The overall effect of Web-based tailored interventions on pain intensity was beneficial, with a small effect size and no significant heterogeneity (SMD –0.21, 95% CI –0.34 to –0.0, P=.003; I^2=29%). Figure 2 shows the forest plot of the SMD in pain intensity. The beneficial pain relief effect was not sustained at medium term (4 RCTs, n=987, SMD –0.08, 95% CI –0.30 to 0.13, P=.45; I^2=48%) and long term (5 RCTs, n=1909, SMD –0.09, 95% CI –0.18 to 0.00, P=.05; I^2=0%). Table 5 provides the details of analyses of the effect on pain intensity for each time point.

Pain-Related Disability

We entered 6 studies (n=953) into the meta-analysis of the short-term effect of Web-based tailored interventions on pain-related disability. The overall effect was significantly beneficial with a small effect size, although heterogeneity was high (SMD –0.38, 95% CI –0.59 to –0.16, P<.001; I^2=58%). Figure 3 shows the forest plot of the SMD in pain-related disability. We entered 3 studies (n=411) into the meta-analysis of the medium-term effect on pain-related disability. The overall effect, although in favor of the experimental group, was not significant (SMD –0.07, 95% CI –0.26 to 0.13, P=.49; I^2=0%). The meta-analysis for long-term effect on pain-related disability could not be conducted due to the lack of assessments available at this time point.

Anxiety, Depression, and Pain Catastrophizing

We found no significant short-term reductions in anxiety (5 RCTs, n=507, SMD –0.08, 95% CI –0.50 to 0.34, P=.70; I^2=79%) or depression (7 RCTs, n=635, SMD –0.33, 95% CI –0.66 to 0.00, P=.05; I^2=73%) with the tailored Web-based intervention compared with standard care or waitlist control. Meta-analytic statistics could not be run for medium-term and long-term effects due to the lack of assessments available at these time points. Meta-analysis could also not be run for pain catastrophizing, which was measured in only 1 study. Multimedia Appendix 1 shows all other forest plots and SMD comparisons with standard care.

Figure 2. Forest plot of the standardized mean difference (95% CI) in pain intensity posttreatment between tailored Web-based interventions and standard care. Black diamond indicates overall treatment effect (tips=95% CI).

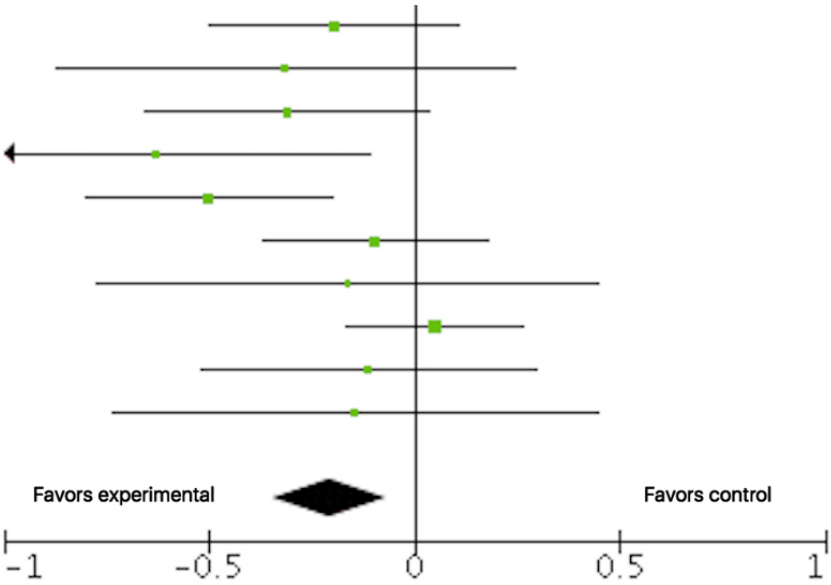


Table 5. Effect of tailored Web-based interventions on pain in the short, medium, and long terms compared with standard care.

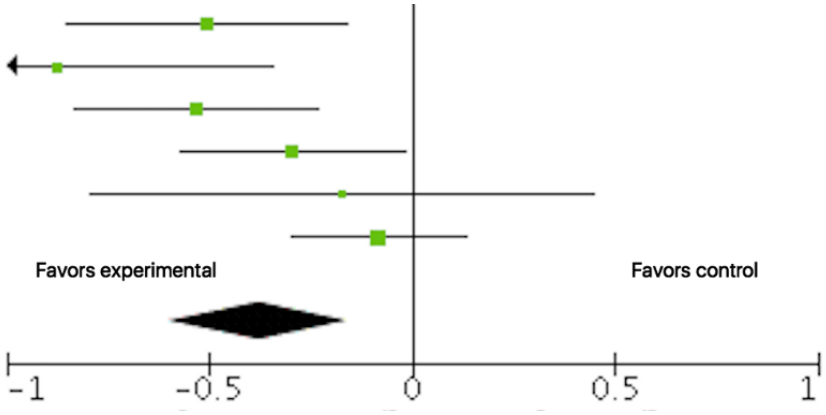
Pain intensity	Study or subgroup				Weight (%)	Standard mean difference IV, random (95% CI)
	Experimental		Standard care			
	Mean (SD)	Total	Mean (SD)	Total		
After completion of intervention[a]						
Bossen [49]	3.5 (4.93)	85	4.5 (5.28)	81	13.0	–0.20 (–0.50 to 0.11)
Buhrman [50]	3.43 (1.68)	22	3.96 (1.63)	29	5.2	–0.32 (–0.87 to 0.24)
Carpenter [51]	5.2 (1.5)	63	5.7 (1.7)	68	11.0	–0.31 (–0.65 to 0.04)
Dear [53]	4.68 (1.7)	30	5.81 (1.85)	30	5.9	–0.63 (–1.15 to –0.11)
Dear [54]	4.86 (1.79)	123	5.71 (1.5)	67	13.2	–0.50 (–0.80 to –0.20)
Krein [55]	5.4 (2.2)	101	5.6 (2)	103	14.8	–0.09 (–0.37 to 0.18)
Moessner [59]	2.17 (1.75)	18	2.54 (2.55)	24	4.4	–0.16 (–0.77 to 0.45)
Moessner [39]	3.74 (2.09)	167	3.64 (2.03)	161	19.1	0.05 (–0.17 to 0.26)
Shigaki [60]	3.68 (2.83)	44	4.02 (3.12)	49	8.7	–0.11 (–0.52 to 0.29)
Strom [61]	2.29 (2.33)	20	2.6 (1.94)	25	4.7	–0.14 (–0.73 to 0.45)
Total (95% CI)		673		637	100.0	–0.21 (–0.34 to –0.07)
At follow-up (<6 months after completion of intervention)[b]						
Buhrman [50]	3.62 (2.04)	22	3.26 (2.16)	29	12.2	0.17 (–0.39 to 0.72)
Lorig [58]	5.86 (2.44)	310	6.34 (2.31)	331	43.5	–0.20 (–0.36 to –0.05)
Moessner [59]	2.67 (1.66)	26	3.46 (2.26)	24	12.0	–0.39 (–0.96 to 0.17)
Moessner [39]	4.18 (2.24)	122	3.97 (2.23)	123	32.3	0.09 (–0.16 to 0.34)
Total (95% CI)		480		507	100.0	–0.08 (–0.30 to 0.13)
At follow-up (>6 months after completion of intervention)[c]						
Bossen [49]	3.5 (4.67)	76	3.8 (4.72)	71	7.7	–0.06 (–0.39 to 0.26)
Lorig [57]	–0.37 (2.72)	354	–0.05 (2.46)	426	40.7	–0.12 (–0.26 to 0.02)
Lorig [58]	5.77 (2.53)	307	6.1 (2.35)	344	34.1	–0.14 (–0.29 to 0.02)
Moessner [39]	4.22 (2.32)	128	4.03 (2.54)	115	12.8	0.08 (–0.17 to 0.33)
Shigaki [60]	4.14 (3.12)	43	3.92 (2.96)	45	4.6	0.07 (–0.35 to 0.49)
Total (95% CI)		908		1001	100.0	–0.09 (–0.18 to 0.00)

[a]Heterogeneity: τ^2=0.01; χ^2_9=12.7 (*P*=.18); I^2=29%. Test for overall effect: Z=2.82 (*P*=.003).

[b]Heterogeneity: τ^2=0.02; χ^2_3=5.8 (*P*=.12); I^2=48%. Test for overall effect: Z=0.76 (*P*=.45).

[c]Heterogeneity: τ^2=0.00; χ^2_4=2.9 (*P*=.58); I^2=0%. Test for overall effect: Z=1.92 (*P*=.05).

Figure 3. Forest plot of the standardized mean difference (95% CI) in pain-related disability posttreatment between tailored Web-based interventions and standard care. Black diamond indicates overall treatment effect (tips=95% CI).



Effects of Tailored Web-Based Interventions Versus Active Control Group

Pain

The tailored Web-based intervention did not relieve pain significantly better than active control in the short term (4 RCTs, n=543, SMD –0.09, 95% CI –0.25 to 0.08, *P*=.32; I^2=0%) or medium term (4 RCTs, n=630, SMD –0.14, 95% CI –0.29 to 0.02, *P*=.09; I^2=0%). We could not analyze the long-term effects on pain because of lack of assessments available at this time point. Table 6 provides the details of analyses of the effect on pain intensity for each time point.

Pain-Related Disability

The tailored Web-based intervention did not relieve pain significantly better than the active control in the short term (2 RCTs, n=426, SMD –0.09, 95% CI –0.28 to 0.10, *P*=.37; I^2=0%) or medium term (2 RCTs, n=411, SMD –0.01, 95% CI –0.20 to 0.19, *P*=.93; I^2=0%). The long-term effect on pain-related disability could not be analyzed due to a lack of assessments available at this time point.

Anxiety, Depression, and Pain Catastrophizing

We entered 3 studies (n=450) into the meta-analysis of the short-term effect on anxiety. The overall effect was not significant (SMD –0.05, 95% CI –0.24 to 0.13, *P*=.56; I^2=0%). In regard to medium-term effect on anxiety, we entered 2 studies (n=411) into the meta-analysis, and the overall effect was not significant (SMD 0.03, 95% CI –0.27 to 0.32, *P*=.87; I^2=55%). Long-term effect on anxiety could not be explored due to a lack of assessments available at this time point.

We entered 3 studies (n=450) into the meta-analysis to analyze the short-term effect on depression. The overall effect was not significant (SMD –0.09, 95% CI –0.28 to 0.09, *P*=.33; I^2=0%). We explored the medium-term effect on depression with 2 studies (n=411). The overall effect was not significant (SMD –0.04, 95% CI –0.30 to 0.21, *P*=.74; I^2=43%). Long-term effects on depression could not be analyzed due again to a lack of assessments at this time point.

The only significant effect of the tailored Web-based intervention was observed for pain catastrophizing in the short term (2 RCTs, n=333, SMD –0.46, 95% CI –0.67 to –0.24, *P*<.001; I^2=0%). Medium-term and long-term effects could not be explored due to lack of assessments available at this time point in this subgroup of studies. Multimedia Appendix 2 shows all forest plots and SMD comparisons with active control groups.

Narrative Review

We could not include 1 study in the meta-analysis because it was the only study targeting acute pain after cardiac surgery [62]. The Web-based tailored intervention was compared with standard care. No effect on pain intensity was recorded. However, less pain interference with breathing and coughing (*P*=.04) was reported by the experimental group, as well as fewer negative pain beliefs and attitudes (*P*=.02).

Funnel Plot Asymmetry and Possible Sources of Explanation

We examined funnel plot asymmetry for the meta-analysis of the effect on pain intensity when comparing tailored Web-based interventions versus standard care posttreatment (short term) solely because other analyses included too few studies (n<10) [64] (see Multimedia Appendix 3 for funnel plot).

Table 6. Effect of tailored Web-based interventions on pain in the short and medium terms compared with active control group.

Pain intensity	Study or subgroup				Weight (%)	Standard mean difference IV, random (95% CI)
	Experimental		Standard care			
	Mean (SD)	Total	Mean (SD)	Total		
After completion of intervention[a]						
Andersson [48]	4 (4.9)	17	3.1 (2.4)	13	5.4	0.22 (–0.51 to 0.94)
Chiauzzi [52]	5.13 (1.95)	95	5.35 (1.94)	104	36.8	–0.11 (–0.39 to 0.17)
Dear [54]	4.86 (1.79)	123	5.2 (1.8)	104	41.7	–0.19 (–0.45 to 0.07)
Kristjánsdóttir [46]	5.41 (2.41)	47	5.06 (2.34)	40	16.0	0.15 (–0.28 to 0.57)
Total (95% CI)		282		261	100.0	–0.09 (–0.25 to 0.08)
At follow-up (<6 months after completion of intervention)[b]						
Chiauzzi [52]	4.78 (2.44)	95	5.18 (2.24)	104	31.6	–0.17 (–0.45 to 0.11)
Dear [54]	4.96 (2)	115	5.02 (1.93)	97	33.6	–0.03 (–0.30 to 0.24)
Kristjánsdóttir [46]	5.19 (2.38)	37	5.85 (2.25)	40	12.2	–0.28 (–0.73 to 0.17)
Leveille [56]	3.3 (2.9)	71	3.8 (3.1)	71	22.6	–0.17 (–0.50 to 0.16)
Total (95% CI)		318		312	100.0	–0.14 (–0.29 to 0.02)

[a]Heterogeneity: τ^2=0.00; χ^2_3=2.5 (*P*=.48); I^2=0%. Test for overall effect: Z=0.99 (*P*=.32).

[b]Heterogeneity: τ^2=0.00; χ^2_3=1.1 (*P*=.78); I^2=0%. Test for overall effect: Z=1.70 (*P*=.09).

We observed some asymmetry at the bottom, possibly reflecting moderate heterogeneity due to the inclusion of small studies. Variations in samples coming from the general population as opposed to outpatients from a clinic could also be a potential explanation.

Discussion

Principal Findings

We explored the efficacy of tailored Web-based interventions for pain management in comparison with standard care or waiting-list controls and active controls. A total of 17 studies met the inclusion criteria. All studies used feedback as 1 of the tailoring mechanisms, and half of them (n=8) used a hybrid format including telephone or face-to-face contact with a therapist. Most of the studies compared the tailored Web-based intervention versus standard or waiting-list control (n=12). Only 1 study concerned acute pain, which we removed from the meta-analysis, resulting in 16 studies available for quantitative assessment. We evaluated 5 different outcomes: pain intensity (primary outcome), pain-related disability, anxiety, depression, and pain catastrophizing. We assessed effects according to 3 time intervals (short term: <1 month; medium term: 1-6 months; and long term: 6-12 months).

Compared with standard care or waiting list, pain intensity (10 RCTs, n=1310) and pain-related disability (6 RCTs, n=953) were improved immediately after the tailored Web-based intervention with small effect sizes (ie, <0.40). No other improvements were observed at follow-up in the medium and long terms. Other systematic reviews and meta-analyses on Web-based CBT [15,18,19], as well as traditional CBT [65], also reported small effects in the reduction of pain posttreatment compared with standard care or waiting list. One meta-analysis showed that small positive effects on pain-related disability were maintained at follow-up (3 months or more) [18]. However, evaluation of effects at follow-up did not distinguish between medium-term and long-term intervals (eg, 3 and 12 months). Moreover, the meta-analysis did not discriminate between active controls and standard or waitlist controls. Although results were in favor of tailored Web-based interventions, we found no statistically significant benefits for anxiety and depression in our study at any of the time points, and the level of heterogeneity was high. Another meta-analysis found a small effect on depression posttreatment compared with standard care. However, that analysis merged both depression and anxiety outcomes [19].

When comparing the active control group, we found no improvements for the primary outcome (pain intensity) or any of the outcomes except for a small effect size on pain catastrophizing (2 RCTs, n=333) immediately after the intervention. The 2 studies included in this analysis were quite different [46,52]. Although they both used content matching, they used different format (hybrid vs not), and a different approach (CBT and motivational vs CBT, acceptance and commitment therapy, and mindfulness). Nonetheless, it is important to underline that the levels of catastrophizing recorded were high. One meta-analysis on Web-based CBT found a small effect on pain catastrophizing posttreatment [19]. However, a high heterogeneity was reported and the comparator was standard care or waiting list. Our results in relation to the active control comparator are in concordance with a previous meta-analysis on traditional psychological therapies for chronic pain, which concluded that there was no evidence of efficacy of CBT and behavioral therapy on pain [65]. It is also noteworthy that meta-analyses conducted in our study with the active control group included very few studies.

Lastly, based on our results and previous results, a general observation can be made pertaining to the effects of Web-based interventions on chronic pain. Regardless of the type of comparison used (ie, usual treatment or active control), effects were redundantly small, which poses a question concerning the adequacy of pain reduction as an outcome. Previous authors [18] have suggested including participants with moderate to severe pain intensity at baseline in order to appreciate the benefits of these interventions. This was true for the only 2 studies in our review that reported moderate effects on pain [53,54]. All participants had a moderate level of pain intensity at baseline. Another avenue could be to focus on outcomes related to the concept of chronic pain acceptance and quality of life [66,67], as most interventions for chronic pain used a CBT approach aimed at reducing disability, depression, and anxiety, not necessarily reducing pain.

Limitations and Future Research

Although our results are consistent with other meta-analyses, a major difference in our results is related to the heterogeneity of intervention approaches that we found in tailored Web-based interventions for pain management as opposed to traditional CBT. Selected studies included interventions profoundly influenced by CBT, but approaches were oftentimes multimodal and varied from wearing a pedometer and physical activity coaching to relaxation and mindfulness. Many of the interventions also included an educational aspect related to management of a specific chronic disease and wellness. The dosage of interventions was also very diverse, ranging from 2 sessions per week for 3 weeks [51] or 5 sessions per week for 4 weeks [46,47] to 12 to 15 weekly sessions [39] or even weekly reminders for 12 months [55]. This variation does not allow for gauging the influence of these interventions. Another observation regarding intervention content is that only half of the interventions used the mechanism of content matching, which is thought to be the essence of tailored approaches [21]. While the only 2 studies that reported moderate effects (ie, <0.50) on pain intensity did not use content matching but rather a hybrid format involving telephone contact with a therapist [53,54], carefully choosing behavioral change techniques that fit the targeted behavior could enhance the effects of these interventions [12,68]. The contribution of this tailoring mechanism is yet to be explored regarding pain management. When the number of studies available allows it, it would be interesting to examine effect according to tailoring mechanisms and dosage.

Attrition and fidelity of intervention delivery are challenges in Web-based interventions. The attrition rate ranged from 5% to 56% with an average of 22%, even though all interventions included some kind of feedback, with half of them using human

interaction. This rate is similar to the rate found in another meta-analysis, underlining that there was no difference in attrition even when participants received more guidance as opposed to no reminders or feedback [19]. Questions remain regarding this issue and the efficacy of these methods. A meta-analysis on tailored Web-based interventions in general found that expert input does not necessarily mean more efficacy [30]. The intervention itself could also be one aspect of the problem. Indeed, interventions for chronic pain require engagement, and some involve frequent weekly activities and are of long duration. Web-based tailored interventions for chronic conditions should consider disease burden and the complexity of accomplishing certain tasks [30]. Although we cannot outline clear guidelines, among studies with higher dropout rates (<20%), interventions could include up to 5 sessions per week [46,47,58] or could last for 12 to 15 weeks [59]. Having a better understanding of which ingredient works for which patient and, consequently, using tailoring to adapt content could lead to more concise and efficient interventions. In fact, it has been observed that pain management interventions with a duration of less than 8 weeks are more effective [69]. Moreover, cost-benefit analyses would definitely be informative, as only small effects have been recorded.

In an effort to decrease heterogeneity in measures and their involved concepts when looking at pain-related disability, we focused on the 2 most commonly used measures: the Roland Morris Disability Questionnaire and the Brief Pain Inventory. This could have influenced our results. Although our study is 1 of the few making a distinction between comparators, we could include only 5 studies in the meta-analysis focusing on active control groups. Our results are in that sense to be interpreted with caution.

Lastly, we found only 1 intervention for acute pain, which highlights the need to explore the avenue of tailored Web-based interventions for this type of pain and for the prevention of chronic pain. The context and settings of acute pain are very different from the chronic spectrum and could lead to different outcomes. Hence, it has been shown that Web-based tailored interventions that are preventive in nature and targeting a general population, not a specific condition, were more successful [30]. Another possible avenue for intervention development is the consideration of sex differences. The majority of participants in the included studies were women, which was observed in other meta-analyses on Web-based interventions for pain management [18,19]. These interventions are tailored to some extent, but none of them have taken into account sex differences in terms of pain experience and coping styles. Although there is still some controversy around pain sensitivity differences, getting a better understanding of women's needs as opposed to men's needs could improve the uptake of interventions. A recent study (n=1371) in the context of a rehabilitation program found that women have better activity level, pain acceptance, and social support, while men report more fear of movement and mood disturbances [70].

Implications and Conclusion

Although several meta-analyses and systematic reviews of Web-based interventions for pain management have been conducted, this is, to our knowledge, the first examining tailored Web-based interventions and using active control groups as a comparator. Tailored Web-based interventions did not prove to be more efficacious than standardized Web-based interventions in terms of pain intensity, pain-related disability, anxiety, and depression. Similar findings to other meta-analyses on Web-based interventions for pain management were generated, meaning that these interventions may have a short-term effect on pain intensity and disability compared with usual care. An interesting finding was that some efficacy was shown in pain catastrophizing compared with active control interventions.

Tailored Web-based interventions are a recent field of research among behavioral change interventions. Considering the diversity of approaches used in tailored Web-based interventions for chronic pain management, their efficacy is yet to be explored. Moreover, the scarcity of tailored Web-based interventions available for acute pain management reflects the multitude of possibilities for intervention development. Feedback was used in all studies but content matching, the most important tailoring ingredient, was used in only half of the studies. More studies would improve our understanding of the efficacy of these interventions, enabling subgroup analyses according to their therapeutic content and their level of tailoring (ie, content matching).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Forest plots and standardized mean differences: standard care.

[PDF File (Adobe PDF File), 116KB - jmir_v19i11e385_app1.pdf]

Multimedia Appendix 2

Forest plots and standardized mean differences: active control.

[PDF File (Adobe PDF File), 127KB - jmir_v19i11e385_app2.pdf]

Multimedia Appendix 3

Funnel plot: tailored Web-based interventions vs standard care (n=10).

[PDF File (Adobe PDF File), 13KB - jmir_v19i11e385_app3.pdf]

References

1. Brennan F, Carr DB, Cousins M. Pain management: a fundamental human right. Anesth Analg 2007 Jul;105(1):205-221. [doi: 10.1213/01.ane.0000268145.52345.55] [Medline: 17578977]
2. Cousins MJ, Brennan F, Carr DB. Pain relief: a universal human right. Pain 2004 Nov;112(1-2):1-4. [doi: 10.1016/j.pain.2004.09.002] [Medline: 15494176]
3. Institute of Medicine of the National Academies Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education and Research. Washington, DC: The National Academies Press; 2011.
4. International Association for the Study of Pain (IASP). Global Year Against Acute Pain 2010-2011. 2010. Acute pain: what is the problem? URL: http://www.iasp-pain.org/AM/Template.cfm?Section=Fact_Sheets3&Template=/CM/ContentDisplay.cfm&ContentID=11783; [accessed 2017-08-30] [WebCite Cache ID 6t6dHGOT4]
5. Nahin RL. Estimates of pain prevalence and severity in adults: United States, 2012. J Pain 2015 Aug;16(8):769-780 [FREE Full text] [doi: 10.1016/j.jpain.2015.05.002] [Medline: 26028573]
6. Fashler SR, Cooper LK, Oosenbrug ED, Burns LC, Razavi S, Goldberg L, et al. Systematic review of multidisciplinary chronic pain treatment facilities. Pain Res Manag 2016;2016:5960987 [FREE Full text] [doi: 10.1155/2016/5960987] [Medline: 27445618]
7. International Association for the Study of Pain (IASP). Pain: Clinical Updates. Washington, DC: IASP Press; 2012. Interdisciplinary Chronic Pain Management: International Perspectives URL: https://www.iasp-pain.org/files/Content/ContentFolders/Publications2/PainClinicalUpdates/Archives/PCU_20-7_web.pdf [accessed 2017-08-30] [WebCite Cache ID 6t6drWgrV]
8. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]
9. Williams DA. Web-based behavioral interventions for the management of chronic pain. Curr Rheumatol Rep 2011 Sep 14;13(6):543-549. [doi: 10.1007/s11926-011-0212-8] [Medline: 21913061]
10. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
11. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
12. 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]
13. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
14. Bender JL, Radhakrishnan A, Diorio C, Englesakis M, Jadad AR. Can pain be managed through the Internet? A systematic review of randomized controlled trials. Pain 2011 Aug;152(8):1740-1750. [doi: 10.1016/j.pain.2011.02.012] [Medline: 21565446]
15. Macea DD, Gajos K, Daglia CYA, Fregni F. The efficacy of Web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis. J Pain 2010 Oct;11(10):917-929. [doi: 10.1016/j.jpain.2010.06.005] [Medline: 20650691]
16. El-Metwally A. Internet-based interventions for pain management: a systematic review of randomised controlled trial (RCTs) conducted from 2010 to 2014. J Public Health Epidemiol 2015 May 31;7(5):170-182. [doi: 10.5897/JPHE2014.0702]
17. Garg S, Garg D, Turin TC, Chowdhury MFU. Web-based interventions for chronic back pain: a systematic review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
18. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (Internet-delivered) for the management of chronic pain in adults. Cochrane Database Syst Rev 2014;2:CD010152. [doi: 10.1002/14651858.CD010152.pub2] [Medline: 24574082]
19. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: a systematic review. Internet Interv 2016 May;4:17-34. [doi: 10.1016/j.invent.2015.12.001]
20. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
21. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]

22. Petty RE, Wegener DT, Fabrigar LR. Attitudes and attitude change. Annu Rev Psychol 1997 Feb;48:609-647. [doi: 10.1146/annurev.psych.48.1.609] [Medline: 9046570]
23. Dijkstra A. Working mechanisms of computer-tailored health education: evidence from smoking cessation. Health Educ Res 2005 Oct;20(5):527-539 [FREE Full text] [doi: 10.1093/her/cyh014] [Medline: 15701665]
24. Petty R, Rucker D, Bizer G, Cacioppo J. The elaboration likelihood model of persuasion. In: Seiter JS, Gass RH, editors. Perspectives on Persuasion, Social Influence, and Compliance Gaining. Boston, MA: Pearson Education; 2004:65-93.
25. Rimer BK, Kreuter MW. Advancing tailored health communication: a persuasion and message effects perspective. J Commun 2006 Aug;56(s1):S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]
26. Petty R, Cacciopo J. Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York, NY: Springer Verlag; 1986.
27. Kreuter M. Tailoring Health Messages: Customizing Communication With Computer Technology. Mahwah, NJ: L Erlbaum; 2000.
28. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Feb;15(1):1-4 [FREE Full text] [Medline: 10788196]
29. Lustria MLA, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
30. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
31. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;349:g7647 [FREE Full text] [Medline: 25555855]
32. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [Medline: 19622552]
33. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
34. Martorella G, Gélinas C, Bérubé M, Boitor M, Fredericks S, LeMay S. The effect of tailored Web-based interventions on pain in adults: a systematic review protocol. Syst Rev 2016 Apr 12;5:59 [FREE Full text] [doi: 10.1186/s13643-016-0233-5] [Medline: 27072140]
35. Merskey H, Bogduk N. Classification of chronic pain. Seattle, WA: IASP Press; 1994. URL: https://www.iasp-pain.org/files/Content/ContentFolders/Publications2/FreeBooks/Classification-of-Chronic-Pain.pdf [accessed 2017-08-30] [WebCite Cache ID 6t6eY1HKX]
36. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003 Dec;106(3):337-345. [Medline: 14659516]
37. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
38. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008 Feb;9(2):105-121. [doi: 10.1016/j.jpain.2007.09.005] [Medline: 18055266]
39. Moessner M, Aufdermauer N, Baier C, Göbel H, Kuhnt O, Neubauer E, et al. [Efficacy of an internet-delivered aftercare program for patients with chronic back pain]. Psychother Psychosom Med Psychol 2014 Feb;64(2):47-53. [doi: 10.1055/s-0033-1351266] [Medline: 24101036]
40. The Cochrane Collaboration. Cochrane Consumers and Communication Group esources for authors: data extraction template. 2016 Dec. URL: http://cccrg.cochrane.org/author-resources [accessed 2017-08-30] [WebCite Cache ID 6t6el2aoJ]
41. Lundh A, Gøtzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Med Res Methodol 2008 Apr 21;8:22 [FREE Full text] [doi: 10.1186/1471-2288-8-22] [Medline: 18426565]
42. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
43. Savović J, Weeks L, Sterne JAC, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev 2014 Apr 15;3:37 [FREE Full text] [doi: 10.1186/2046-4053-3-37] [Medline: 24731537]
44. The Cochrane Collaboration. RevMan: computer program version 5.3. Copenhagen, Denmark: The Nordic Cochrane Centre; 2014. URL: http://community.cochrane.org/tools/review-production-tools/revman-5 [accessed 2017-08-30] [WebCite Cache ID 6t6es54oI]

45. Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, European Palliative Care Research Collaborative (EPCRC). Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011 Jun;41(6):1073-1093. [doi: 10.1016/j.jpainsymman.2010.08.016] [Medline: 21621130]
46. Kristjánsdóttir Ó, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
47. Kristjánsdóttir Ó, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain. Part 2: 11-month follow-up results of a randomized trial. J Med Internet Res 2013 Mar 28;15(3):e72 [FREE Full text] [doi: 10.2196/jmir.2442] [Medline: 23538392]
48. Andersson G, Lundström P, Ström L. Internet-based treatment of headache: does telephone contact add anything? Headache 2003 Apr;43(4):353-361. [Medline: 12656706]
49. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
50. Buhrman M, Fältenhag S, Ström L, Andersson G. Controlled trial of Internet-based treatment with telephone support for chronic back pain. Pain 2004 Oct;111(3):368-377. [doi: 10.1016/j.pain.2004.07.021] [Medline: 15363881]
51. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
52. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
53. Dear BF, Titov N, Perry KN, Johnston L, Wootton BM, Terides MD, et al. The Pain Course: a randomised controlled trial of a clinician-guided Internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. Pain 2013 Jun;154(6):942-950. [doi: 10.1016/j.pain.2013.03.005] [Medline: 23688830]
54. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The Pain Course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. Pain 2015 Oct;156(10):1920-1935 [FREE Full text] [doi: 10.1097/j.pain.0000000000000251] [Medline: 26039902]
55. Krein SL, Kadri R, Hughes M, Kerr EA, Piette JD, Holleman R, et al. Pedometer-based internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. J Med Internet Res 2013 Aug 19;15(8):e181 [FREE Full text] [doi: 10.2196/jmir.2605] [Medline: 23969029]
56. Leveille SG, Huang A, Tsai SB, Allen M, Weingart SN, Iezzoni LI. Health coaching via an internet portal for primary care patients with chronic conditions: a randomized controlled trial. Med Care 2009 Jan;47(1):41-47. [doi: 10.1097/MLR.0b013e3181844dd0] [Medline: 19106729]
57. Lorig KR, Ritter PL, Laurent DD, Plant K. Internet-based chronic disease self-management: a randomized trial. Med Care 2006 Nov;44(11):964-971. [doi: 10.1097/01.mlr.0000233678.80203.c1] [Medline: 17063127]
58. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
59. Moessner M, Schiltenwolf M, Neubauer E. Internet-based aftercare for patients with back pain-a pilot study. Telemed J E Health 2012 Jul;18(6):413-419. [doi: 10.1089/tmj.2011.0221] [Medline: 22676398]
60. Shigaki CL, Smarr KL, Siva C, Ge B, Musser D, Johnson R. RAHelp: an online intervention for individuals with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2013 Oct;65(10):1573-1581. [doi: 10.1002/acr.22042] [Medline: 23666599]
61. Ström L, Pettersson R, Andersson G. A controlled trial of self-help treatment of recurrent headache conducted via the Internet. J Consult Clin Psychol 2000 Aug;68(4):722-727. [Medline: 10965647]
62. Martorella G, Côté J, Racine M, Choinière M. Web-based nursing intervention for self-management of pain after cardiac surgery: pilot randomized controlled trial. J Med Internet Res 2012;14(6):e177 [FREE Full text] [doi: 10.2196/jmir.2070] [Medline: 23241361]
63. Sterne J, Egger M, Moher D. Addressing reporting biases. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Chichester, UK: John Wiley & Sons, Ltd; 2008.
64. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002. [Medline: 21784880]
65. Williams AC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012;11:CD007407. [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
66. van Boekel RLM, Vissers KCP, van der Sande R, Bronkhorst E, Lerou JGC, Steegers MAH. Moving beyond pain scores: multidimensional pain assessment is essential for adequate pain management after surgery. PLoS One 2017;12(5):e0177345 [FREE Full text] [doi: 10.1371/journal.pone.0177345] [Medline: 28489926]

67. McCracken LM, Vowles KE. Acceptance and commitment therapy and mindfulness for chronic pain: model, process, and progress. Am Psychol 2014;69(2):178-187. [doi: 10.1037/a0035623] [Medline: 24547803]
68. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
69. Carnes D, Homer KE, Miles CL, Pincus T, Underwood M, Rahman A, et al. Effective delivery styles and content for self-management interventions for chronic musculoskeletal pain: a systematic literature review. Clin J Pain 2012 May;28(4):344-354. [doi: 10.1097/AJP.0b013e31822ed2f3] [Medline: 22001667]
70. Rovner GS, Sunnerhagen KS, Björkdahl A, Gerdle B, Börsbo B, Johansson F, et al. Chronic pain and sex-differences; women accept and move, while men feel blue. PLoS One 2017;12(4):e0175737 [FREE Full text] [doi: 10.1371/journal.pone.0175737] [Medline: 28441403]

Abbreviations

CBT: cognitive behavioral therapy
NRS: numeric rating scale
RCT: randomized controlled trial
SMD: standardized mean difference
VAS: visual analog scale

Edited by G Eysenbach; submitted 30.08.17; peer-reviewed by Y Tang, H Eide, M Linnett; comments to author 27.09.17; revised version received 10.10.17; accepted 11.10.17; published 10.11.17.

Please cite as:
Martorella G, Boitor M, Berube M, Fredericks S, Le May S, Gélinas C
Tailored Web-Based Interventions for Pain: Systematic Review and Meta-Analysis
J Med Internet Res 2017;19(11):e385
URL: http://www.jmir.org/2017/11/e385/
doi:10.2196/jmir.8826
PMID:29127076

Original Paper

Effectiveness of a Web 2.0 Intervention to Increase Physical Activity in Real-World Settings: Randomized Ecological Trial

Corneel Vandelanotte[1], PhD; Gregory S Kolt[2], PhD; Cristina M Caperchione[3], PhD; Trevor N Savage[4], MA; Richard R Rosenkranz[5], PhD; Anthony J Maeder[6], PhD; Anetta Van Itallie[1], MA; Rhys Tague[2], PhD; Christopher Oldmeadow[7], PhD; W Kerry Mummery[8], PhD; Mitch J Duncan[7], PhD

[1]Central Queensland University, Rockhampton, Australia
[2]Western Sydney University, Sydney, Australia
[3]University of British Columbia, Kelowna, BC, Canada
[4]School of Allied Health Sciences, Griffith University, Southport, Australia
[5]Kansas State University, Manhattan, KS, United States
[6]Flinders University, Adelaide, Australia
[7]University of Newcastle, Newcastle, Australia
[8]University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Corneel Vandelanotte, PhD
Central Queensland University
Bruce Highway
Rockhampton, 4700
Australia
Phone: 61 749232183
Email: c.vandelanotte@cqu.edu.au

Abstract

Background: The translation of Web-based physical activity intervention research into the real world is lacking and becoming increasingly important.

Objective: To compare usage and effectiveness, in real-world settings, of a traditional Web 1.0 Web-based physical activity intervention, providing limited interactivity, to a Web 2.0 Web-based physical activity intervention that includes interactive features, such as social networking (ie, status updates, online “friends,” and personalized profile pages), blogs, and Google Maps mash-ups.

Methods: Adults spontaneously signing up for the freely available 10,000 Steps website were randomized to the 10,000 Steps website (Web 1.0) or the newly developed WALK 2.0 website (Web 2.0). Physical activity (Active Australia Survey), quality of life (RAND 36), and body mass index (BMI) were assessed at baseline, 3 months, and 12 months. Website usage was measured continuously. Analyses of covariance were used to assess change over time in continuous outcome measures. Multiple imputation was used to deal with missing data.

Results: A total of 1328 participants completed baseline assessments. Only 3-month outcomes (224 completers) were analyzed due to high attrition at 12 months (77 completers). Web 2.0 group participants increased physical activity by 92.8 minutes per week more than those in the Web 1.0 group (95% CI 28.8-156.8; P=.005); their BMI values also decreased more (–1.03 kg/m2, 95% CI –1.65 to -0.41; P=.001). For quality of life, only the physical functioning domain score significantly improved more in the Web 2.0 group (3.6, 95% CI 1.7-5.5; P<.001). The time between the first and last visit to the website (3.57 vs 2.22 weeks; P<.001) and the mean number of days the website was visited (9.02 vs 5.71 days; P=.002) were significantly greater in the Web 2.0 group compared to the Web 1.0 group. The difference in time-to-nonusage attrition was not statistically significant between groups (Hazard Ratio=0.97, 95% CI 0.86-1.09; P=.59). Only 21.99% (292/1328) of participants (n=292 summed for both groups) were still using either website after 2 weeks and 6.55% (87/1328) were using either website after 10 weeks.

Conclusions: The website that provided more interactive and social features was more effective in improving physical activity in real-world conditions. While the Web 2.0 website was visited significantly more, both groups nevertheless displayed high nonusage attrition and low intervention engagement. More research is needed to examine the external validity and generalizability of Web-based physical activity interventions.

Trial Registration: Australian New Zealand Clinical Trials Registry: ACTRN12611000253909; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=336588&isReview=true (Archived by WebCite at http://www.webcitation.org/6ufzw2HxD)

(J Med Internet Res 2017;19(11):e390) doi:10.2196/jmir.8484

KEYWORDS

Internet; online; Web based; behavioral intervention; external validity; pragmatic trial

Introduction

Given low population levels of physical activity and high associated physical and mental burden of disease caused by inactive lifestyles [1], there is a need for effective physical activity interventions that can reach large populations at low cost [2]. In this context, research into the effectiveness of Web-based interventions has become popular [3], as large and increasingly diverse populations can be reached without geographical limitations by using the Internet [4]. While literature reviews and meta-analyses point to the short-term effectiveness of Web-based physical activity interventions [5,6], they also highlight that there is a lack of evidence for long-term behavior change, as well as a lack of knowledge about what are the most effective intervention components [7]. The lack of evidence for long-term behavior change has often been attributed to low levels of participant engagement and retention, due to examining websites that are static in nature, that lack social support elements, and that provide limited opportunity for interactivity or information exchange [5].

Websites with more dynamic, interactive, user-focused features, also referred to as *second generation* or *Web 2.0* features, are now commonly used and include social networking, blogs, wikis, podcasts, and mash-ups [8]. They provide users with the opportunity to directly generate, modify, and share information [9]. Few physical activity studies have examined the effectiveness of Web 2.0 features [8,10,11], which may enhance engagement with the intervention and, in turn, lead to long-term behavior change. A review by Maher et al, however, indicated that the use of online social networks in behavior change trials was only modestly effective [10]. In this context, it has been argued that randomized controlled trials (RCTs) are not the most appropriate research design to truly examine the effectiveness of Web 2.0 features [12].

It has been asserted that the highly controlled nature of RCTs, which aim to minimize impact of selection bias, confounding factors, and contamination, stifle the dynamic, spontaneous, viral nature of Web 2.0 features [12]. While RCTs are an essential component of the research process, complementary approaches with high external validity and generalizability are also essential. For example, if one is not able to invite friends to join an online social network due to RCT-related restrictions, the social network is unlikely to be as functional and effective as it would be in real-world circumstances [13]. As such, there is a need for alternative and ecologically valid research designs that evaluate Web-based interventions in real-life conditions in order to advance the science in this area [14].

The translation and dissemination of Web-based physical activity intervention research into the real world is lacking and becoming increasingly important. Therefore, the main aim of this study was to compare the physical activity behavior of individuals using a traditional Web 1.0 physical activity website to those using an innovative Web 2.0 physical activity website in real-world settings. The study also aimed to assess the effectiveness of Web 2.0 features to engage and retain individuals to a physical activity promotion website, as well as examine differences in quality of life and body mass index (BMI) between intervention groups. The primary hypothesis was that participants in the Web 2.0 condition would display higher levels of physical activity at 3 and 12 months, compared to those in the Web 1.0 condition. The secondary hypotheses were that, in the Web 2.0 condition, there would be higher website engagement and retention as well as improvements in quality of life and BMI when compared to the Web 1.0 condition at 3 and 12 months.

Methods

Overview

As the detailed protocol for this study has been published elsewhere [13], only summary information will be provided here (see Multimedia Appendix 1 for additional screenshots of the intervention). This study is the second phase of the substantive WALK 2.0 project and builds on an earlier RCT [15,16], which rigorously tested the efficacy of the interventions described here [11]. The Western Sydney University Human Research Ethics Committee granted ethical approval for this study (H8767). This trial has been registered at the Australian New Zealand Clinical Trials Registry (ACTRN12611000253909).

Recruitment, Procedures, Study Design, and Participants

Adults 18 years of age or older spontaneously signing up for the freely available and Web-based 10,000 Steps program [17], which attracts over a 1000 new members per month [18], were asked during the registration process whether they wanted to participate in a research study from November 2012 to June 2014. The 10,000 Steps project has been funded since 2001 through Queensland Health, one of the Australian State Ministries of Health. The project is well known through media and marketing events in Australia, particularly in Queensland, with over 70% program awareness in population-based surveys. If potential participants agreed to participate in research, they received more information about the study, were screened online for eligibility, provided informed consent, and completed a brief baseline survey. Using a computer-generated algorithm, they

were randomized to receive access to one of two intervention websites: a Web 1.0 intervention, which was the 10,000 Steps website they were originally signing up for, or a Web 2.0 intervention, which was the newly developed WALK 2.0 website. For technical reasons, participants were randomized before completing the baseline measures, however, they only gained access to intervention materials after completing the baseline assessment. Follow-up outcomes were assessed 3 and 12 months postbaseline using online questionnaires; participants were invited by email and received up to three reminders. All actions, from study invitation to completion, were fully automated with no interaction from the research team at any point. The research team also provided no instructions as to how the interventions should be used and there was no predefined intervention duration. However, even though the aim of reaching 10,000 steps a day was implicit, participants were provided with the Australian Physical Activity Guidelines and the websites were designed to encourage self-monitoring and interaction on a daily basis for as long as possible—participants had an option to receive a daily reminder to use the websites. No pedometers were provided. Exclusion criteria were the following: being under 18 years of age; seeking to participate in a 10,000 Steps Workplace Challenge; having been a participant in the WALK 2.0 RCT; and having a medical condition that prevents them from increasing physical activity, assessed through the Physical Activity Readiness Questionnaire [19].

Interventions

Web 1.0 Intervention

Participants allocated to the Web 1.0 group were given access to the existing 10,000 Steps website. This website was originally developed to promote the community-based 10,000 Steps Australia project [20,21]. The website includes features that support individual self-monitoring (eg, step log) and communication exchange (eg, discussion forums), and provides access to a library with educational resources (eg, benefits of activity). Participants were able to log steps and/or type and duration of other physical activities. Participants also had the ability to share stories, ask questions, or make comments in the discussion forum.

Web 2.0 Intervention

Participants allocated to the Web 2.0 group were given access to a newly developed website, WALK 2.0, that provided content and functionality similar to the Web 1.0 condition (eg, step log and library); however, this website was supplemented with Web 2.0 features that included annotation, messaging, and group-publishing tools implemented in a stand-alone social networking setting. Participants in the Web 2.0 group also had access to self-monitoring features and educational resources, however, these had advanced functionalities that provided greater interactivity and participatory communication between users (eg, status updates, internal emails, inviting "friends," and personalized profile pages). Participants could upload content to their own profile page, share this information with others, and invite individuals who were not study participants to become their "friends" and use the website. Participants could also connect with Facebook (eg, post their step total for the day on their Facebook wall), but the Web 2.0 website was completely separate from Facebook.

Measures

Demographics

Participants' gender, age, educational level (school education, trade/diploma, or higher education), employment status (full time, part time/casual, or other), occupation (professional white collar, blue collar, or other), weekly household income (<Aus $1000, Aus $1000-$1999, Aus $2000-$5000, or no response), Internet self-confidence (low or high), height (cm), and weight (kg) were assessed. Self-reported BMI was calculated as weight (kg) over height squared (m^2) and categorized as normal weight (≤24.99 kg/m^2), overweight (25-29.99 kg/m^2), and obese (≥30 kg/m^2).

Physical Activity

The Active Australia Survey was used to measure self-reported physical activity [22]. This instrument provides an estimate of total weekly minutes of physical activity by summing total minutes of walking for transport and recreation, moderate-intensity physical activity, and vigorous-intensity physical activity—multiplied by 2 to account for the higher energy expenditure per time unit—during the previous week. The Active Australia Survey has acceptable test-retest reliability and validity in Australian adults [23,24] and has been demonstrated to be sensitive enough to detect change over time [25]. The Active Australia Survey was used to determine the following: total physical activity per week and whether participants were engaging in sufficient physical activity—a minimum of 150 minutes of moderate-to-vigorous physical activity per week accumulated over five or more sessions.

Quality of Life

The RAND 36 Short Form Survey was used to assess quality of life in eight health-related categories: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions [26,27]. All items were scored from 0 to 100, with a high score representing a more favorable health state. Items in each category were then averaged together to create eight subscale scores. The RAND 36 has been validated in Australian populations [28].

Website Engagement and Retention

Website usage statistics for both websites were continuously measured using Google Analytics (eg, time on site) and data were extracted directly from the website databases (eg, step entry information). These measures were only examined from baseline to 3 months (first 12 weeks), due to the low survey completion rate at 12 months. The total and average number of website visits were assessed, as well as the time between the first and last visit. The total and average number of days with a step entry and step entry comments were assessed, including the time between first and last step entry. Nonusage attrition was defined as not having visited the website and/or logged steps for at least two consecutive weeks [11,18].

Website Usability

The System Usability Scale (SUS) was used to assess website usability [29]. This scale is a 10-item survey, scored on a 5-point scale of strength of agreement, with good reliability and concurrent validity [30]. Final scores can range from 0 to 100, where higher scores indicate better usability. Self-reported use and usefulness for different features on both websites is also reported. Participants were asked about usefulness on a 5-point scale; the proportion of participants who thought the feature was "useful" or "very useful" is reported.

Sample Size

The trial was powered to detect a 4% between-group difference in the prevalence of sufficient physical activity, as defined by the Australian Physical Activity Guidelines, between the Web 1.0 and Web 2.0 groups. To achieve this aim with 80% power and an alpha level of .05, a minimum of 1034 participants per group were needed [13].

Statistical Analyses

All statistical analyses were conducted using SAS version 9.4 (SAS Institute Inc). Differences between participants with complete and missing data were compared using *t* tests or Pearson chi-square tests. Analysis of covariance (ANCOVA) was used to test for differences between treatment groups at 3 months in physical activity, quality of life, and BMI; baseline physical activity, BMI, quality of life levels, and confounding variables were included in the models as covariates. Logistic regression was used to estimate between-group differences in the proportion of participants who achieved sufficient physical activity. Results are presented both for those with complete data at baseline and 3 months (*completer analyses*), as well as those with missing data, following intention-to-treat principles. Multiple imputation was applied to deal with missing data, under the missing-at-random assumption, using the chained equations method. Rubin's method was used to pool the treatment effects using 25 imputed datasets, as the fraction of missing data was high [31]. To analyze between-group differences in website engagement and retention, *t* tests were used. A proportional hazards regression model was used to estimate between-group differences in time from randomization to nonusage; Kaplan-Meier estimates of the proportion remaining active (the *survival distribution*) are also presented [32]. Due to very small participant numbers, 12-month data were not included in any of the analyses. The significance level was set at $P<.05$.

Results

Participants

A participant flowchart is provided in Figure 1. After automatically screening out ineligible people, 10,673 people were invited and 3480 indicated an interest in participating. After eligibility checks, providing informed consent, and website registration, 1328 people completed all baseline measures. Out of 1328 participants, 224 (16.87%) completed the 3-month assessment and 77 (5.80%) completed the 12-month assessment. Table 1 presents participant demographics. At baseline, the majority of participants were female (1095/1328, 82.45%), were 44 years of age or under (818/1328, 61.60%), were overweight or obese (849/1328, 63.93%), had a higher education (693/1328, 52.18%), were full-time employed (771/1328, 58.06%), had a professional or white-collar job (879/1328, 66.19%), and participated in sufficient physical activity (757/1328, 57.00%). There were no significant between-group differences in participant characteristics at baseline; however, several between-group differences were observed among those who completed the 3-month assessment and those who did not. There was significantly greater retention among those who were Web 1.0 group participants, male, aged 45 years or older, and not obese, as well as those having a higher education, professional occupation, and higher income.

Physical Activity

The physical activity outcomes are presented in Table 2. In the intention-to-treat analysis, a significant difference between groups was found: participants randomized to the Web 2.0 group increased physical activity by 92.8 minutes per week more compared to those in the Web 1.0 group ($P=.005$; Cohen *d* effect size=0.29). In the completer analysis, participants in the Web 2.0 group increased physical activity by 56.6 minutes per week more compared to those in the Web 1.0 group, however, this difference was not significant ($P=.20$; Cohen *d* effect size=0.24).

At baseline, 57.00% (757/1328) of participants in both groups engaged in sufficient physical activity. At the 3-month time point, 77% (62/80) of Web 2.0 participants and 71.5% (103/144) of Web 1.0 participants engaged in sufficient physical activity. A significant between-group difference in favor of the Web 2.0 group was observed in the intention-to-treat analysis (Relative Risk [RR]=1.11, 95% CI 1.01-1.21; $t_{64.5}=2.19$, $P=.03$), but not in the completer analysis (RR=1.03, 95% CI 0.89-1.18; Z=0.36, $P=.71$).

Quality of Life

There was no effect of the intervention on most quality of life variables (see Table 2), except for the physical functioning domain: a significant improvement (3.6 units) was observed in the Web 2.0 group compared to the Web 1.0 group ($P<.001$).

Body Mass Index

BMI reduced over time in both groups (see Table 2) and significant between-group differences were observed for both the intention-to-treat analysis (in favor of the Web 2.0 group: change in BMI=-1.03 kg/m^2, $P=.002$) and the completer analysis (in favor of the Web 1.0 group: change in BMI=-0.58 kg/m^2, $P=.002$).

Figure 1. Participant flowchart. PAR-Q: Physical Activity Readiness Questionnaire.

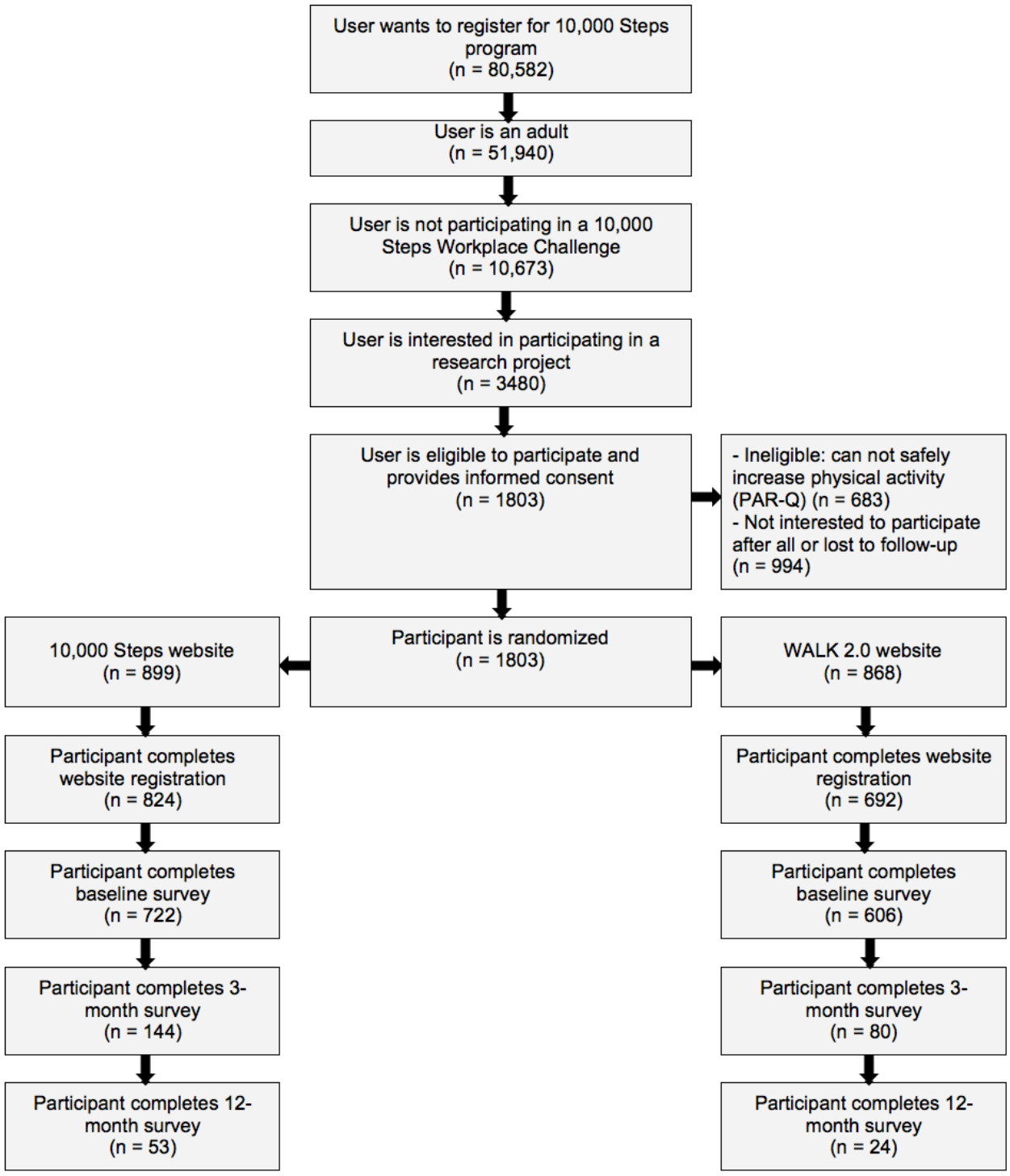


Table 1. Participant demographic characteristics at baseline by group and for those with complete or missing data.

Demographic characteristics	Total (n=1328)	Web 1.0: 10,000 Steps (n=722)	Web 2.0: WALK 2.0 (n=606)	Completer at 3 months (n=224)	Missing at 3 months (n=1104)	*P* value[a]
Group, n (%)						
10,000 Steps	722 (54.36)	722 (100)	0 (0)	144 (64.3)	578 (52.35)	.001
WALK 2.0	606 (45.63)	0 (0)	606 (100)	80 (35.7)	526 (47.73)	
Gender, n (%)						
Male	233 (17.54)	117 (16.2)	116 (19.1)	54 (24.1)	179 (16.21)	.006
Female	1095 (82.45)	605 (83.8)	490 (80.9)	170 (75.9)	925 (83.78)	
Age (years), n (%)						
18-34	503 (37.87)	278 (38.5)	225 (42.1)	52 (23.2)	451 (40.85)	<.001
35-44	315 (23.71)	163 (23.6)	152 (25.1)	45 (20.1)	270 (24.45)	
45-54	310 (23.34)	174 (24.1)	136 (22.4)	67 (29.9)	243 (22.01)	
55-64	164 (12.34)	83 (11.5)	81 (13.4)	43 (19.2)	121 (10.96)	
65 and over	36 (2.71)	24 (3.3)	12 (2.0)	17 (7.6)	19 (1.72)	
Internet self-confidence, n (%)						
Low	38 (2.86)	24 (3.3)	14 (2.3)	6 (2.7)	32 (2.89)	.86
High	1289 (97.06)	698 (96.7)	591 (97.5)	218 (97.3)	1071 (97.01)	
Body mass index (kg/m^2), n (%)						
Normal weight (≤24.99)	438 (32.98)	239 (33.1)	199 (32.8)	83 (37.1)	355 (32.15)	.008
Overweight (25-29.99)	376 (28.31)	202 (28.0)	174 (23.8)	75 (33.5)	301 (27.26)	
Obese (≥30)	473 (35.61)	259 (35.9)	214 (35.3)	61 (27.2)	412 (37.31)	
Highest education level, n (%)						
Higher education	698 (51.88)	362 (50.1)	336 (55.4)	135 (60.3)	563 (50.99)	.03
Trade/diploma	428 (32.22)	240 (33.2)	188 (31.0)	64 (28.6)	364 (32.97)	
School education	201 (15.13)	120 (16.6)	81 (13.4)	25 (11.2)	176 (15.94)	
Employment, n (%)						
Full time	771 (58.05)	425 (58.9)	346 (57.1)	140 (61.9)	631 (57.15)	.35
Part time/casual	263 (19.80)	130 (18.0)	133 (21.9)	39 (17.4)	224 (20.28)	
Other	293 (22.06)	167 (23.1)	126 (20.8)	45 (20.1)	248 (22.46)	
Occupation, n (%)						
Professional	540 (40.66)	298 (40.0)	242 (39.9)	111 (49.6)	429 (38.85)	.04
White collar	339 (25.52)	177 (24.5)	162 (26.7)	45 (20.1)	294 (26.63)	
Blue collar	37 (2.78)	19 (2.6)	18 (3.0)	7 (3.1)	30 (2.71)	
Other	119 (8.96)	61 (8.4)	58 (9.6)	16 (7.1)	103 (9.32)	
No response	293 (22.06)	167 (23.1)	126 (20.8)	45 (20.1)	248 (22.46)	
Weekly household income (Aus $), n (%)						
<$1000	361 (27.18)	193 (26.7)	168 (27.7)	40 (18.3)	321 (29.07)	.001
$1000-$1999	372 (28.01)	207 (28.7)	165 (27.2)	74 (33.0)	298 (26.99)	
$2000-$5000	324 (24.39)	173 (24.0)	151 (24.9)	67 (29.9)	257 (23.27)	
No response	270 (20.33)	149 (20.6)	121 (20.0)	43 (19.2)	227 (20.56)	
Total weekly minutes of physical activity, mean (SD)	338 (348)	341 (353)	334 (342)	306 (270)	344 (361)	.13

Demographic characteristics	Total (n=1328)	Web 1.0: 10,000 Steps (n=722)	Web 2.0: WALK 2.0 (n=606)	Completer at 3 months (n=224)	Missing at 3 months (n=1104)	*P* value[a]
Level of physical activity, n (%)						
Sufficient	757 (57.00)	410 (56.8)	347 (57.3)	141 (62.9)	616 (55.79)	.13
Insufficient	496 (37.34)	274 (38.0)	222 (36.6)	71 (32.0)	425 (38.49)	
No reported activity	75 (5.64)	38 (5.3)	37 (6.1)	12 (5.4)	63 (5.70)	

[a]*P* values compare those who completed or were missing at the 3-month time point. Statistical significance is represented by $P<.05$.

Table 2. Changes in outcome variables for intention-to-treat (n=1328) and completer (n=224) analyses.

Outcome variables	Web 1.0: 10,000 Steps		Web 2.0: WALK 2.0		Difference in between-group changes (95% CI)	Z or *t*[a]	*P* value[b]
	Baseline, mean (SD)	3 months, mean (SD/SE[c])	Baseline, mean (SD)	3 months, mean (SD/SE[c])			
Total physical activity (min/week) [d]							
Intention-to-treat	341.9 (353.8)	381.7 (16.6)	334.0 (342.2)	473.9 (26.4)	+92.78 (28.78- 156.77)	2.91	.005
Completers	341.9 (353.8)	385.2 (269.5)	334.0 (342.2)	464.5 (359.0)	+56.65 (-30.35-143.64)	1.28	.20
Quality of life [e]							
Physical functioning							
Intention-to-treat	88.8 (15.5)	89.1 (0.8)	88.5 (16.2)	92.7 (0.5)	+3.58 (1.66-5.49)	3.72	.001
Completers	88.8 (15.5)	89.6 (14.5)	88.5 (16.2)	92.9 (7.7)	+3.79 (0.68-6.91)	2.39	.02
Role limitations due to physical health problems							
Intention-to-treat	83.8 (29.3)	83.8 (1.6)	84.2 (29.0)	84.2 (1.7)	+0.38 (-4.85-5.62)	0.15	.88
Completers	83.8 (29.3)	84.2 (29.0)	84.2 (29.0)	85.3 (28.0)	-0.14 (-8.18-7.90)	0.04	.97
Role limitation due to personal or emotional problems							
Intention-to-treat	74.3 (36.9)	80.6 (2.1)	76.2 (36.8)	84.0 (1.8)	+3.31 (-2.59-9.22)	1.12	.27
Completers	74.3 (36.9)	80.1 (35.4)	76.2 (36.8)	85.0 (28.8)	+5.83 (-2.68-14.34)	1.34	.18
Energy/fatigue							
Intention-to-treat	50.0 (20.7)	57.2 (1.1)	51.4 (20.4)	59.5 (1.4)	+2.28 (-1.49-6.06)	1.21	.23
Completers	50.0 (20.7)	57.8 (21.2)	51.4 (20.4)	60.2 (19.0)	+1.55 (-2.84-5.94)	0.69	.49
Emotional well-being							
Intention-to-treat	69.3 (19.2)	76.4 (0.9)	70.2 (19.1)	78.0 (1.0)	+1.55 (-1.25-4.34)	1.11	.27
Completers	69.3 (19.2)	76.1 (15.6)	70.2 (19.1)	78.1 (16.1)	+1.16 (-2.39-4.71)	0.64	.52
Social functioning							
Intention-to-treat	79.5 (24.0)	84.0 (1.3)	80.7 (23.0)	86.4 (1.2)	+2.28 (-1.48- 6.04)	0.22	.23
Completers	79.5 (24.0)	84.4 (20.4)	80.7 (23.0)	87.2 (18.4)	+3.10 (-1.81-8.01)	1.24	.22
Bodily pain							
Intention-to-treat	79.1 (20.7)	78.9 (1.2)	80.6 (20.0)	79.4 (1.5)	+0.32 (-3.29-3.92)	0.18	.86
Completers	79.1 (20.7)	79.4 (19.7)	80.6 (20.0)	79.9 (19.0)	+0.18 (-4.82-5.19)	0.07	.94
General health perceptions							
Intention-to-treat	62.7 (20.1)	66.6 (1.1)	64.3 (20.4)	69.5 (1.3)	+2.64 (-1.10-6.38)	1.42	.16
Completers	62.7 (20.1)	67.4 (17.0)	64.3 (20.4)	69.7 (18.1)	+1.60 (-1.79-4.99)	0.93	.35
Body mass index (kg/m^2) [f]							
Intention-to-treat	28.6 (6.4)	28.3 (0.2)	28.8 (6.5)	27.3 (0.3)	-1.03 (-1.65-0.41)	3.33	.001
Completers	28.6 (6.4)	26.8 (5.6)	28.8 (6.5)	27.2 (4.5)	-0.58 (-0.95-0.21)	3.08	.002

[a]*t* values are presented for imputed data; Z values (Wald chi-square test: Z=sqrt[chisq]) are presented for completers data.

[b]Statistical significance is represented by $P<.05$.

[c]Standard error (SE) of the mean values are presented for imputed data (intention-to-treat); standard deviation (SD) values are presented for completers data.

[d]In addition to controlling for baseline physical activity, the analyses were adjusted for gender, age, body mass index, and education.

[e]In addition to controlling for the baseline value of the outcome variable, the analyses were adjusted for baseline physical activity, gender, age, body mass index, and education.

[f]In addition to controlling for baseline body mass index, the analyses were adjusted for baseline physical activity, gender, age, and education.

Website Engagement, Retention, and Usability

Differences in terms of website usage are shown in Table 3. Participants used the websites on average 3 minutes per week and logged steps for approximately 10 days. There were no significant between-group differences with regard to time on site, entering steps without comments, and the SUS. However, total number of visits (9.0 [19.3] vs 5.7 [17.9], *P*=.002) and average number of visits (0.7 [1.6] vs 0.4 [1.4], *P*=.002) were significantly higher in the Web 2.0 group, as well as the time between first and last visit (3.6 [3.7] vs 2.2 [2.9], *P*<.001). Total days (0.7 [5.9] vs 0.2 [1.5], *P*=.03) and average days (0.05 [0.45] vs 0.01 [0.10], *P*=.02) with a step entry comment were also significantly higher in the Web 2.0 group. A Kaplan-Meier survival plot shows the proportion of participants that remained using the website for each week of the study (see Figure 2). Only 21.99% (292/1328) of participants were still using either website after 2 weeks (n=292 summed for both groups) and 6.55% (87/1328) after 10 weeks. The between-group difference in time-to-nonusage attrition was not statistically significant (Hazard Ratio=0.97, 95% CI 0.86-1.09; *P*=.57). Self-reported use and usefulness of the features available on both websites are reported in Table 4. More participants in the Web 2.0 group used features that were present, or similar, on both websites compared to the Web 1.0 group. Web 2.0 participants also rated the usefulness of these features more highly compared to Web 1.0 participants. Many participants in both groups did not use some of the interactive features or indicate that they were very useful.

Table 3. Website engagement, retention, and usability between weeks 1 and 12.

Engagement, retention, and usability metrics	Web 1.0: 10,000 Steps (n=565), mean (SD)	Web 2.0: WALK 2.0 (n=697), mean (SD)	*t*	*P* value[a]
Time on site (seconds per week)	195 (464)	179 (678)	-0.482	.63
Total number of visits	5.71 (17.95)	9.02 (19.34)	3.140	.002
Average number of visits per week	0.46 (1.48)	0.73 (1.58)	3.096	.002
Time between the first and last visit (weeks)	2.22 (2.92)	3.57 (3.75)	7.187	<.001
Total number of days with a step entry	9.97 (21.44)	9.10 (20.88)	-0.731	.47
Average number of days with a step entry per week	0.78 (1.71)	0.71 (1.67)	-0.693	.49
Time between the first and last step entry (weeks)	1.72 (3.43)	1.74 (3.43)	0.070	.95
Total number of days with a step entry comment	0.19 (1.46)	0.70 (5.99)	2.169	.03
Average number of days with a step entry comment per week	0.01 (0.10)	0.05 (0.45)	2.256	.02
Average time for nonusage attrition to occur (weeks)	1.55 (3.16)	1.54 (3.11)	-0.530	.96
Systems Usability Score	61.73 (10.76)	62.53 (11.12)	0.520	.60

[a]Statistical significance is represented by *P*<.05.

Figure 2. Number of participants engaged with the study and intervention at different time points.

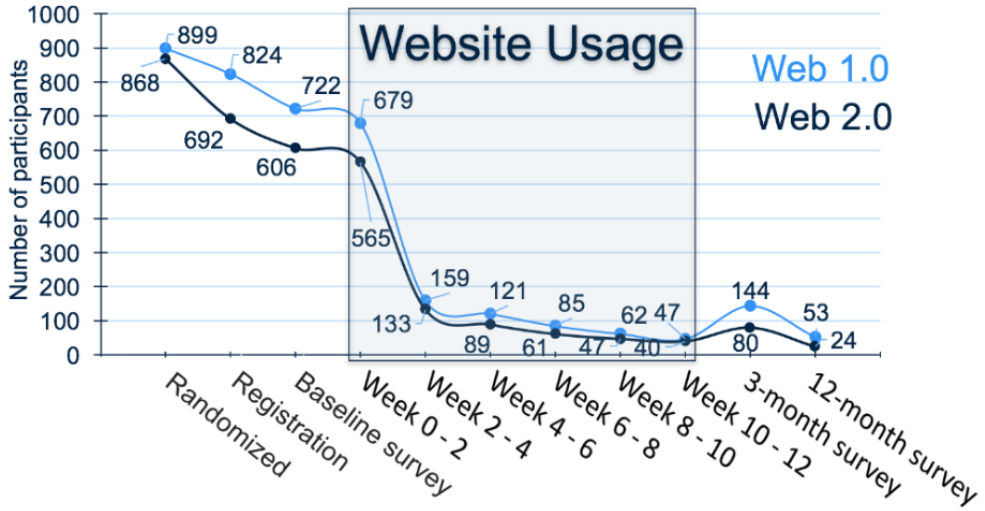


Table 4. Self-reported use and usefulness of the features available on both websites.

Website features[a]	Web 1.0 (n=144)		Web 2.0 (n=80)	
	Did not use, n (%)	Useful or very useful, n (%)	Did not use, n (%)	Useful or very useful, n (%)
Step entry tool	16 (14.5)	91 (63.1)	6 (8)	63 (79)
Ability to view your step progress	22 (15.2)	89 (61.6)	7 (9)	66 (83)
Articles in the library	51 (35.5)	35 (24.6)	24 (30)	22 (28)
Ability to set goals	32 (22.5)	59 (41.3)	14 (17)	48 (61)
Ability to have and view progress of walking buddies (Web 1.0)/friends (Web 2.0)	58 (40.6)	29 (20.3)	29 (37)	20 (25)
Discussion forum	69 (47.8)	14 (9.4)	29 (37)	13 (16)
Group-based challenges	65 (45.7)	17 (11.6)	21 (26)	17 (21)
Monthly individual challenges	45 (31.2)	42 (29.0)	N/A[b]	N/A
Ability to share your story	72 (50.0)	10 (7.2)	N/A	N/A
Ability to read others' stories	67 (46.4)	23 (15.9)	N/A	N/A
Ability to like and comment on friends' updates	N/A	N/A	32 (40)	18 (22)
Ability to send messages to other users	N/A	N/A	33 (41)	17 (21)
Profile page to provide updates	N/A	N/A	31 (38)	19 (24)
Ability to have your own walking blog	N/A	N/A	36 (45)	15 (18)
Ability to read friends' blog posts	N/A	N/A	31 (38)	19 (24)
Google Maps walking tool	N/A	N/A	40 (50)	15 (18)
Ability to connect with Facebook	N/A	N/A	38 (47)	12 (15)

[a]Some features were present on only one of the two intervention websites; hence, no data are available for those features on the other website.

[b]N/A: not applicable.

Discussion

Principal Findings

The aim of this study was to compare physical activity behavior of individuals using a traditional Web 1.0 physical activity website with those using a more interactive and social Web 2.0 physical activity website in real-world settings. The primary hypothesis was confirmed: those in the Web 2.0 group did display significantly higher levels of physical activity at 3 months compared to those in the Web 1.0 group. The secondary hypotheses were partially confirmed: BMI significantly decreased, engagement was significantly higher in some variables (eg, website visits), and quality of life significantly improved in one variable (ie, physical functioning) in the Web 2.0 group compared to the Web 1.0 group at 3 months. However, there were no between-group differences for several other engagement variables (eg, logging steps and nonusage attrition) and most quality-of-life variables. This study was the first to demonstrate the importance of using Web 2.0 features in Web-based physical activity interventions in a real-life setting. These outcomes are strengthened by not finding between-group differences in website usability (ie, SUS score), indicating that outcomes were not influenced by factors such as *user friendliness*. Interestingly, the use and usability of features that were present, or similar, on both websites were higher in the Web 2.0 website compared to the Web 1.0 website. While, from this research, it is not possible to explain this difference given the similar SUS scores, it may in part explain why the Web 2.0 website performed better on several behavioral and engagement outcomes.

While there is an abundance of studies examining the effectiveness of physical activity promotion websites in controlled conditions [33], few studies have examined the use and effectiveness of physical activity websites in natural and real-life conditions. To our knowledge, only one other study has examined people who spontaneously signed up to an online physical activity intervention and where researchers had no direct contact with study participants: all processes were completed automatically [34]. Wanner et al conducted an RCT with two groups of participants that were actively recruited to be randomized to either a control group or an interactive, computer-tailored, physical activity website; however, they included a third group of spontaneous users from the same website. Significantly larger increases in physical activity were observed in the spontaneous users when compared to the actively recruited groups [34]. That study's outcomes were similar to ours when comparing findings from our RCT with the ecological trial: physical activity increases reported in this ecological trial (+140 for Web 2.0 group; +40 for Web 1.0 group) are much higher than those observed in the RCT (+45.5 for Web 2.0 group; –1.0 for Web 1.0 group) [11]. The large differences observed in these studies indicate that it is not adequate to rely

only on RCT outcomes when preparing Web-based physical activity interventions for dissemination and translation.

Poor engagement with Web-based interventions has often been reported; with regard to that aspect, this study is not unique [33,35]. The ecological trial and the RCT [11], however, were remarkably different in terms of engagement and retention outcomes. While the Web 2.0 group also had better engagement outcomes in the RCT, the overall engagement of participants was much lower in the ecological trial. For example, nonusage attrition in the RCT occurred on average after 35.5 and 25.5 weeks in the Web 2.0 and Web 1.0 groups, respectively, whereas it occurred after a mere 1.5 weeks for both groups in the ecological trial. The RCT website was visited, on average, 3.6 and 1.6 times per week for the Web 2.0 and Web 1.0 groups, respectively, but only 0.7 and 0.5 times per week, respectively, in the ecological trial. While participants in both trials were exposed to identical websites, there are factors that encourage engagement in RCTs, and factors that discourage engagement in ecological trials. In the RCT, participants were required to complete multiple face-to-face visits and they received phone calls, a pedometer, and multiple small financial incentives [11]. These strategies increased participant accountability and resulted in a 60% retention rate at 18 months, which is comparable to other studies in this field [6]. In order to keep the intervention implementation as natural as possible in the ecological trial, no such strategies, other than email reminders, were applied. However, in the ecological trial the process of asking 10,000 Steps visitors to participate in research and obtaining informed consent and baseline assessments may have deterred many participants, given the attrition at each step of the process (shown in Figure 1). These necessary steps may have been detrimental to their ongoing participation, as is demonstrated in a previous analysis of 16,948 spontaneous 10,000 Steps users who did not need to go through multiple screenings to be included in the analyses [18]. Among these 10,000 Steps users, nonusage attrition occurred after 4 weeks and the website was visited 2.4 times per week; these engagement outcomes are remarkably higher than those among ecological trial participants [18]. Finally, in another study, Wanner et al compared spontaneous users with trial participants and also reported large engagement differences: nonusage attrition occurred at 41 weeks in trial participants, but it was zero days in spontaneous users [36]. Collectively, these outcomes indicate that caution should be taken when interpreting engagement outcomes from both RCTs and ecological trials, as neither may be a good reflection of how participants engage with websites in reality. The outcomes further suggest that improved participant retention strategies are required that work well in ecologically valid circumstances [37].

The significant reduction of BMI in favor of the Web 2.0 group was surprising, as the intervention did not focus on weight loss, and the 3-month time frame is reasonably short. However, previous Web-based interventions have demonstrated weight loss, both in interventions focused on either weight loss or physical activity [38,39]. Further, few studies have examined how Web-based physical activity interventions can improve health-related quality of life. A meta-analysis of the impact of non-Web-based physical activity interventions on quality of life observed that the interventions only significantly changed the physical functioning domain of quality of life, and not other domains [40]. This is consistent with the outcomes of this study.

Women dominated participation in this study (82%), however, this was very comparable to other studies. For example, Anderson-Bill et al [41] conducted an online walking program and reported that 83% of participants were female and 75% of spontaneous website users were female in the study by Wanner et al [36]. Lower male participation is also commonly observed in RCTs [5,6] and this study demonstrates that the difficulty in attracting men to health behavior change interventions may even be greater in ecologically valid circumstances. Several significant differences were observed between those who participated at 3 months and those who did not. There was less attrition among those randomized to the Web 1.0 group (10,000 Steps). As all participants were originally signing up to participate in the 10,000 Steps program, it is plausible that many of those who were randomized to the Web 2.0 group (WALK 2.0) were disappointed with their allocation and dropped out for that reason. It is also possible that the increased complexity and interactivity of the Web 2.0 website resulted in higher dropout rates, though this is less likely as there was no significant difference in the SUS score between the websites. Older participants and men were less likely to drop out; this is similar to the study by Wanner et al, who found that among their spontaneous users, men and those of increasing age were more likely to repeatedly use their website [36]. It is also in line with a Web-based physical activity intervention study that found older participants spent more time on the website and changed behavior more than younger participants [42]. Finally, it is not surprising to see lower dropout rates among those with higher education, higher income, and a professional occupation, as it has been widely reported that people with a higher socioeconomic status are more amenable to participating in health behavior change interventions [43,44].

Strengths and Limitations

The major strengths of this study were the innovative ecological randomized nature of the study and the comparison of Web 1.0 and Web 2.0 intervention features. The large nonusage and study attrition, however, was an important limitation, making it problematic to analyze 12-month outcomes. This limitation, though, could be considered as a finding that is of interest, as it is a reflection of how Web-based interventions are being used in ecologically valid circumstances; it is not, per se, a reflection of poor study methodology. That said, several methodological limitations are inherent to ecological trials, such as the lack of a true control group and having to resort to less intrusive (ie, self-report) measures to assess outcomes. These methodological concerns can, however, be alleviated by comparing the study outcomes to those of an RCT that applied more rigorous methods [11]. Furthermore, all engagement and retention measures were assessed objectively through the use of website usage statistics. However, it was a limitation of this study that the interventions were only accessible via websites, as the technology landscape has rapidly changed since the study inception and it is now very common to also use mobile phone apps, which are more convenient to access than websites [18].

Conclusions

Web 2.0 intervention features appear to be more effective in increasing physical activity, decreasing BMI, and improving physical functioning (ie, quality-of-life domain) compared to Web 1.0 features in ecologically valid circumstances. While the social and interactive Web 2.0 features increased engagement compared to a traditional Web 1.0 website, website usage was low regardless, especially when compared to our previous RCT [11]. As such, more research is needed to increase our understanding of how people engage with Web-based interventions, both in controlled and ecologically valid circumstances, and how closely this engagement is related to actual behavior change (ie, what is the dose-response relationship?) [45]. Finally, the findings of this study are remarkable in how different they are from the findings observed in controlled conditions in terms of attrition, website usage, and behavior changes observed. This is important, because information obtained through RCTs may not translate well to real-world conditions. Yet almost all of our knowledge in this field is derived from RCTs and this has a major influence on how Web-based interventions are designed, what interventions are deemed effective, and what interventions are being disseminated and implemented. As such, there is an urgent need for more ecological trials and implementation studies.

Acknowledgments

This trial was funded by the National Health and Medical Research Council of Australia (project grant number 589903). The funder did not have any role in the study other than to provide funding. CV (fellowship ID 100427) and MJD (fellowship ID 100029) are each supported by a Future Leader Fellowship from the National Heart Foundation of Australia.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Conference presentation slides and screenshots.

[PPTX File, 1MB - jmir_v19i11e390_app1.pptx]

Multimedia Appendix 2

CONSORT-EHEALTH checklist (V.1.6.1).

[PDF File (Adobe PDF File), 601KB - jmir_v19i11e390_app2.pdf]

References

1. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Lancet Physical Activity Series 2 Executive Committee. The economic burden of physical inactivity: A global analysis of major non-communicable diseases. Lancet 2016 Sep 24;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
2. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC, Lancet Physical Activity Series 2 Executive Committee. Scaling up physical activity interventions worldwide: Stepping up to larger and smarter approaches to get people moving. Lancet 2016 Sep 24;388(10051):1337-1348. [doi: 10.1016/S0140-6736(16)30728-0] [Medline: 27475273]
3. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
4. Australian Communications and Media Authority. Communications Report 2015-16. Canberra, Australia: Commonwealth of Australia; 2016. URL: http://www.acma.gov.au/theACMA/Library/researchacma/Research-reports/communications-report-2015-16 [accessed 2017-07-12] [WebCite Cache ID 6rtOUM8sT]
5. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions: A review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
6. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of Internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
7. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to Internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
8. Chou WS, Prestin A, Lyons C, Wen K. Web 2.0 for health promotion: Reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-e18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]

9. O'Reilly T. O'Reilly. 2005 Sep 30. What is Web 2.0: Design patterns and business models for the next generation of software URL: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [accessed 2017-07-12] [WebCite Cache ID 6rtOtuUab]
10. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
11. Kolt GS, Rosenkranz RR, Vandelanotte C, Caperchione CM, Maeder AJ, Tague R, et al. Using Web 2.0 applications to promote health-related physical activity: Findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Jan 03:1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]
12. Vandelanotte C, Maher CA. Why we need more than just randomized controlled trials to establish the effectiveness of online social networks for health behavior change. Am J Health Promot 2015;30(2):74-76. [doi: 10.4278/ajhp.141204-CIT-605] [Medline: 26517586]
13. Caperchione CM, Kolt GS, Savage TN, Rosenkranz RR, Maeder AJ, Vandelanotte C, et al. WALK 2.0: Examining the effectiveness of Web 2.0 features to increase physical activity in a 'real world' setting: An ecological trial. BMJ Open 2014;4(10):e006374 [FREE Full text] [doi: 10.1136/bmjopen-2014-006374] [Medline: 25304191]
14. Victora CG, Habicht J, Bryce J. Evidence-based public health: Moving beyond randomized trials. Am J Public Health 2004 Mar;94(3):400-405. [Medline: 14998803]
15. Kolt GS, Rosenkranz RR, Savage TN, Maeder AJ, Vandelanotte C, Duncan MJ, et al. WALK 2.0 - Using Web 2.0 applications to promote health-related physical activity: A randomised controlled trial protocol. BMC Public Health 2013 May 03;13:436 [FREE Full text] [doi: 10.1186/1471-2458-13-436] [Medline: 23642010]
16. Caperchione C, Duncan M, Rosenkranz R, Vandelanotte C, Van Itallie A, Savage T, et al. Recruitment, screening, and baseline participant characteristics in the WALK 2.0 study: A randomized controlled trial using Web 2.0 applications to promote physical activity. Contemp Clin Trials Commun 2016 Apr;2:25-33 [FREE Full text] [doi: 10.1016/j.conctc.2015.12.004]
17. 10,000 Steps. URL: https://www.10000steps.org.au/ [accessed 2017-10-26] [WebCite Cache ID 6uVKSWV7g]
18. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: The case of 10,000 Steps Australia. J Med Internet Res 2015;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
19. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
20. Brown W, Mummery W, Eakin E, Schofield G. 10,000 Steps Rockhampton: Evaluation of a whole community approach to improving population levels of physical activity. J Phys Act Health 2006 Jan;3(1):1-14 [FREE Full text] [doi: 10.1123/jpah.3.1.1]
21. Brown W, Eakin E, Mummery W, Trost S. 10,000 Steps Rockhampton: Establishing a multi-strategy physical activity promotion project in a community. Health Promot J Austr 2003;14:95-100 ISSN: 1036-1073 [FREE Full text] [doi: 10.1071/HE03095]
22. Australian Institute of Health and Welfare. The Active Australia Survey: A Guide and Manual for Implementation, Analysis and Reporting. Canberra, Australia: Australian Institute of Health and Welfare; 2003 Apr 24. URL: http://www.aihw.gov.au/publication-detail/?id=6442467449 [accessed 2017-07-12] [WebCite Cache ID 6rtQe6Aq9]
23. Brown WJ, Trost SG, Bauman A, Mummery K, Owen N. Test-retest reliability of four physical activity measures used in population surveys. J Sci Med Sport 2004 Jun;7(2):205-215. [Medline: 15362316]
24. Brown WJ, Burton NW, Marshall AL, Miller YD. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. Aust N Z J Public Health 2008 Dec;32(6):535-541. [doi: 10.1111/j.1753-6405.2008.00305.x] [Medline: 19076744]
25. Reeves MM, Marshall AL, Owen N, Winkler EAH, Eakin EG. Measuring physical activity change in broad-reach intervention trials. J Phys Act Health 2010 Mar;7(2):194-202. [Medline: 20484758]
26. VanderZee KI, Sanderman R, Heyink JW. Psychometric qualities of the RAND 36-Item Health Survey 1.0: A multidimensional measure of general health status. Int J Behav Med 1996;3(2):104-122. [doi: 10.1207/s15327558ijbm0302_2] [Medline: 16250758]
27. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ 1993 Oct;2(3):217-227. [Medline: 8275167]
28. McCallum J. The SF-36 in an Australian sample: Validating a new, generic health status measure. Aust J Public Health 1995 Apr;19(2):160-166. [Medline: 7786942]
29. Brooke J. SUS: A "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McCelland IL, editors. Usability Evaluation In Industry. London, UK: Taylor & Francis; 1996:189-194.
30. Lewis JR, Sauro J. The factor structure of the System Usability Scale. In: Kurosu M, editor. Human Centered Design. Berlin, Germany: Springer Berlin Heidelberg; 2009:94-103.
31. Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66(5):688-701 [FREE Full text] [doi: 10.1037/h0037350]

32. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958 Jun;53(282):457-481 [FREE Full text] [doi: 10.2307/2281868]
33. Kohl LFM, Crutzen R, de Vries N. Online prevention aimed at lifestyle behaviors: A systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
34. Wanner M, Martin-Diener E, Braun-Fahrländer C, Bauer G, Martin BW. Effectiveness of active-online, an individually tailored physical activity intervention, in a real-life setting: Randomized controlled trial. J Med Internet Res 2009;11(3):e23 [FREE Full text] [doi: 10.2196/jmir.1179] [Medline: 19666456]
35. 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]
36. Wanner M, Martin-Diener E, Bauer G, Braun-Fahrländer C, Martin BW. Comparison of trial participants and open access users of a Web-based physical activity intervention regarding adherence, attrition, and repeated participation. J Med Internet Res 2010;12(1):e3 [FREE Full text] [doi: 10.2196/jmir.1361] [Medline: 20147006]
37. Robinson KA, Dinglas VD, Sukrithan V, Yalamanchilli R, Mendez-Tellez PA, Dennison-Himmelfarb C, et al. Updated systematic review identifies substantial number of retention strategies: Using more strategies retains more study participants. J Clin Epidemiol 2015 Dec;68(12):1481-1487 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.04.013] [Medline: 26186981]
38. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of Web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: A systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
39. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
40. Kelley G, Kelley K, Hootman J, Jones D. Exercise and health-related quality of life in older community-dwelling adults. J Appl Gerontol 2009 Jan 26;28(3):369-394 [FREE Full text] [doi: 10.1177/0733464808327456]
41. Anderson-Bill ES, Winett RA, Wojcik JR. Social cognitive determinants of nutrition and physical activity among Web-health users enrolling in an online intervention: The influence of social support, self-efficacy, outcome expectations, and self-regulation. J Med Internet Res 2011;13(1):e28 [FREE Full text] [doi: 10.2196/jmir.1551] [Medline: 21441100]
42. Ammann R, Vandelanotte C, de Vries H, Mummery WK. Can a website-delivered computer-tailored physical activity intervention be acceptable, usable, and effective for older people? Health Educ Behav 2013 Apr;40(2):160-170. [doi: 10.1177/1090198112461791] [Medline: 23077157]
43. Nagler RH, Ramanadhan S, Minsky S, Viswanath K. Recruitment and retention for community-based eHealth interventions with populations of low socioeconomic position: Strategies and challenges. J Commun 2013 Feb 1;63(1):201-220 [FREE Full text] [doi: 10.1111/jcom.12008] [Medline: 23439871]
44. Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in physical activity: Review and update. Med Sci Sports Exerc 2002 Dec;34(12):1996-2001. [doi: 10.1249/01.MSS.0000038974.76900.92] [Medline: 12471307]
45. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]

Abbreviations

ANCOVA: analysis of covariance
BMI: body mass index
N/A: not applicable
RCT: randomized controlled trial
RR: Relative Risk
SUS: System Usability Scale

Edited by I Weber; submitted 17.07.17; peer-reviewed by J Kernot, D Reinwand; comments to author 20.08.17; revised version received 02.09.17; accepted 04.09.17; published 13.11.17.

Please cite as:
Vandelanotte C, Kolt GS, Caperchione CM, Savage TN, Rosenkranz RR, Maeder AJ, Van Itallie A, Tague R, Oldmeadow C, Mummery WK, Duncan MJ
Effectiveness of a Web 2.0 Intervention to Increase Physical Activity in Real-World Settings: Randomized Ecological Trial
J Med Internet Res 2017;19(11):e390
URL: http://www.jmir.org/2017/11/e390/
doi:10.2196/jmir.8484
PMID:29133282

Review

Self-Guided Web-Based Interventions: Scoping Review on User Needs and the Potential of Embodied Conversational Agents to Address Them

Mark R Scholten[1], MA, MBA; Saskia M Kelders[1,2], PhD; Julia EWC Van Gemert-Pijnen[1], PhD

[1]Centre for eHealth & 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:
Mark R Scholten, MA, MBA
Centre for eHealth & Wellbeing Research
Department of Psychology, Health and Technology
University of Twente
Drienerlolaan 5
Enschede, 7522 NB
Netherlands
Phone: 31 651409623
Email: m.r.scholten@utwente.nl

Abstract

Background: Web-based mental health interventions have evolved from innovative prototypes to evidence-based and clinically applied solutions for mental diseases such as depression and anxiety. Open-access, self-guided types of these solutions hold the promise of reaching and treating a large population at a reasonable cost. However, a considerable factor that currently hinders the effectiveness of these self-guided Web-based interventions is the high level of nonadherence. The absence of a human caregiver apparently has a negative effect on user adherence. It is unknown to what extent this human support can be handed over to the technology of the intervention to mitigate this negative effect.

Objective: The first objective of this paper was to explore what is known in literature about what support a user needs to stay motivated and engaged in an electronic health (eHealth) intervention that requires repeated use. The second objective was to explore the current potential of embodied conversational agents (ECAs) to provide this support.

Methods: This study reviews and interprets the available literature on (1) support within eHealth interventions that require repeated use and (2) the potential of ECAs by means of a scoping review. The rationale for choosing a scoping review is that the subject is broad, diverse, and largely unexplored. Themes for (1) and (2) were proposed based on grounded theory and mapped on each other to find relationships.

Results: The results of the first part of this study suggest the presence of user needs that largely remain implicit and unaddressed. These support needs can be categorized as task-related support and emotion-related support. The results of the second part of this study suggest that ECAs are capable of engaging and motivating users of information technology applications in the domains of learning and behavioral change. Longitudinal studies must be conducted to determine under what circumstances ECAs can create and maintain a productive user relationship. Mapping the user needs on the ECAs' capabilities suggests that different kinds of ECAs may provide different solutions for improving the adherence levels.

Conclusions: Autonomous ECAs that do not respond to a user's expressed emotion in real time but take on empathic roles may be sufficient to motivate users to some extent. It is unclear whether those types of ECAs are competent enough and create sufficient believability among users to address the user's deeper needs for support and empathy. Responsive ECAs may offer a better solution. However, at present, most of these ECAs have difficulties to assess a user's emotional state in real time during an open dialogue. By conducting future research with relationship theory–based ECAs, the added value of ECAs toward user needs can be better understood.

(J Med Internet Res 2017;19(11):e383) doi:10.2196/jmir.7351

KEYWORDS

eHealth; review; embodied conversational agent; human computer interaction; clinical psychology; health behavior; Web-based intervention; adherence; intelligent tutoring system; ITS

Introduction

Meta-analyses have demonstrated that Web-based interventions for mental health have become reasonably successful treatments against common mental health problems such as depression and anxiety [1-3]. However, it is a consistent finding that human-supported Web-based therapeutic interventions outperform self-guided interventions [4] (in which there is no support from a human). The mere remote presence of a human being delivering informational support, emotional support, or a therapeutic service results in significantly higher effect sizes [5]. In addition, human-supported interventions achieve higher rates of adherence; more participants use the intervention as intended, for example, by completing all the lessons of an intervention [1,3,6]. Nonadherence is an important issue in Web-based interventions for mental health [7] and becomes an even bigger problem when evidence-based therapies are deployed as free-to-access self-guided Web-based therapeutic interventions [8]. In these interventions, adherence, defined as the percentage of users who complete all lessons, falls to a level as low as 1.0% [7] or even 0.5% [8].

The higher rates of adherence in human-supported interventions can be explained in favor of therapists who do an effective job in motivating clients during their change process [5]. However, positive effects of electronic interventions have also been found by using features such as reminders and tailored advice [9]. Interestingly enough, Talbot [10] concludes in her meta-study that the involvement of a professional support provider, a therapist, is not key. Instead, a minimal level of nonguiding human contact is key. Irrespective of whether this type of contact is provided by a layperson or a professional, it has equally large positive effects on intervention adherence. Moreover, scheduling support on one's own can already have an effect on treatment effectiveness. A telephone contact, scheduled at the start of reading a self-help book, yields surprisingly large completion rates and treatment outcomes [11]. This poses the question: what support is needed to achieve higher rates of adherence and effectiveness? A study of Cavanagh and Millings [12] provides evidence of built-in "common factors," such as generating hope, empathy, warmth, collaboration, and feedback, that increase the effectiveness of interventions. However, there is no generally accepted definition of these "common factors." The urgency of support is expressed by the statement of Kreijns et al [13] who declare that the reason that digital learning environments fail is because of socioemotional processes being "ignored, neglected, or forgotten." As Web-based health interventions share many characteristics with digital learning environments, it is a fair assumption that the same socioemotional processes play a role and should be subject to study in relation to adherence.

The following challenge is how these socioemotional processes could be handled within Web-based health interventions. As suggested by Bickmore [14], putting an embodied conversational agent (ECA; also called relational agent) as an adjunct to a self-care application can be a means to support users. An ECA, according to Bickmore, is *a computational artifact designed to build long-term socioemotional relationships with users*. Within the context of health care, Bickmore [14] suggests that the ECA-user relationship can contribute to trust and therapeutic alliance for the purpose of *enhancing adherence to self-care treatment regimens*.

Altogether, ECAs hold the promise that they can bring in social, emotional, and relational elements to the user interface. It is, however, less clear to what extent ECAs can (1) truly handle *expressed needs* of users of electronic health (eHealth) interventions and (2) provide user stimulation that will truly pay out in terms of *user adherence to eHealth interventions*. This was the rationale for us to conduct the research as described within this paper. Therefore, the aim of this study was to structurally analyze existing literature to extract (1) user needs pertaining to eHealth interventions and (2) the capabilities of ECAs to fulfill these needs within the larger objective of enhancing user adherence.

Methods

Study Design

This study was performed by means of structured data collection within the Web of Science and Scopus databases. The scoping review was chosen as research method. A scoping review aims to map the existing literature in a field of interest in terms of the volume, nature, and characteristics of the primary research [15]. The rationale for choosing a scoping review for the subject of this paper is that research on Web-based interventions forms a large and diverse body of literature. Within this literature, the role of support and its relationship to user motivation are barely explored and poorly understood. This is equally the case for system support provided by ECAs within, for example, social learning contexts [16]. To the best of our knowledge, no studies were conducted that systematically aimed to match user needs for Web-based interventions to ECA capabilities to find potential solutions for low adherence to the interventions. Having said that, seminal studies (eg, [17]), have suggested and partly demonstrated that ECAs have the potential to stimulate and motivate users, which ultimately may have a positive effect on intervention adherence—which underscores the importance of this study.

This study is divided into two parts:

> Part 1: a scoping review of meta-studies on user support in Web-based interventions. The focus of this review was on generic user support needs, irrespective of the intervention type and type of disorder.
>
> Part 2: a scoping review of the opportunities of embodied conversational agents to deliver support within Web-based interventions for health or learning.

Research Question Part 1

Is there a set of generic user support needs that are currently not sufficiently addressed within eHealth interventions requiring repeated use that may result in a lower user experience and therefore lower user adherence?

Search Strategy Part 1: Meta-Studies on Support in Web-Based Interventions

The Scopus database was searched with a combination of the concepts "support," "Web-based intervention," and "review." For each of the concepts, multiple keywords were used (see Multimedia Appendix 1). Textbox 1 provides the inclusion and exclusion criteria.

The search resulted in 93 studies. On the basis of our inclusion and exclusion criteria, we selected 18 studies. By checking the references of these selected studies, we found another 4 relevant papers. Finally, 22 papers were included. See Figure 1 for the selection process.

Textbox 1. Inclusion and exclusion criteria.

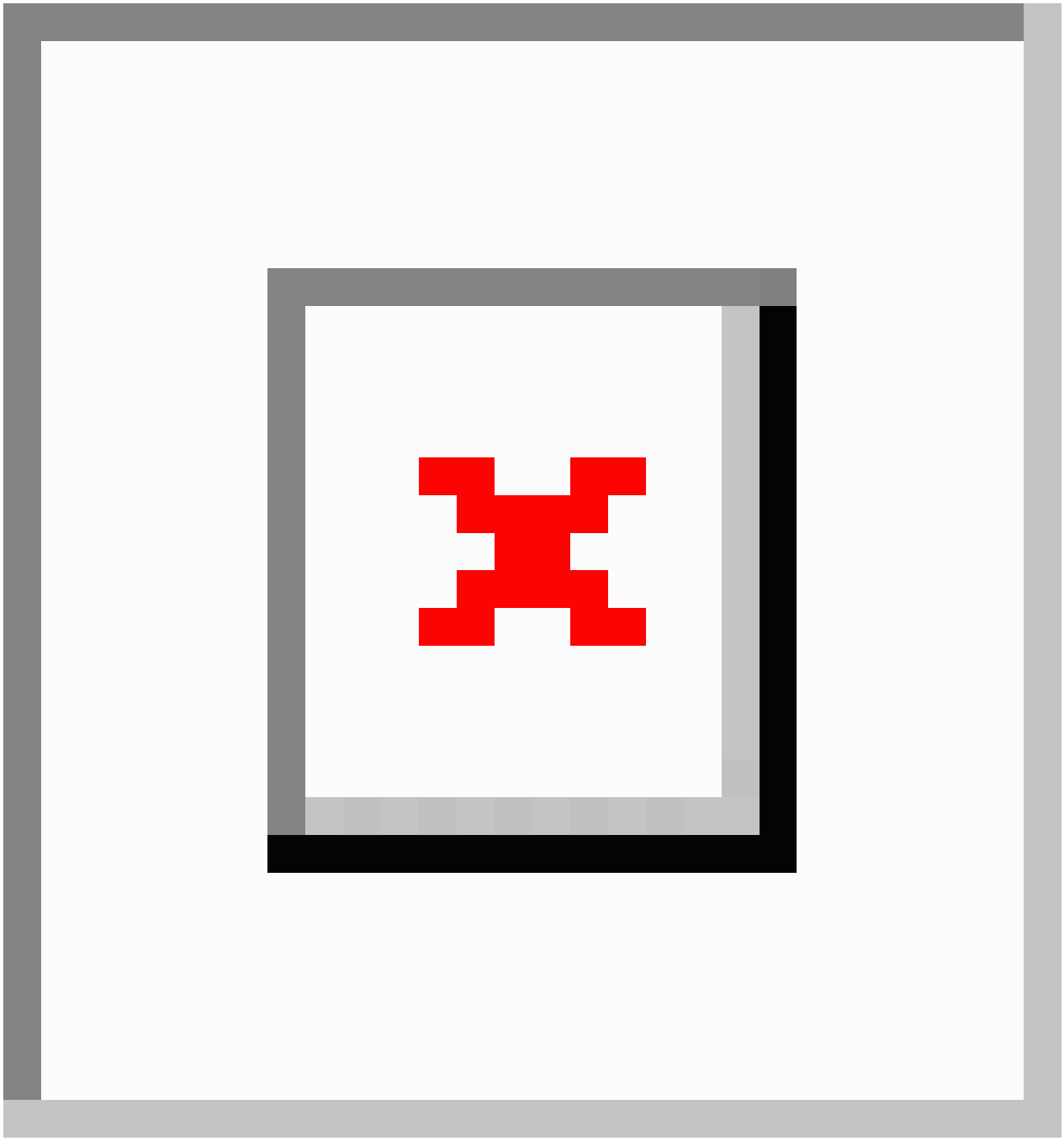
Inclusion criteria were as follows:

- Papers had to address a Web-based intervention for a mental or physical disorder in which support was the subject of the study
- Papers had to review multiple interventions/studies or present ideas based on literature or an earlier study

Exclusion criteria were as follows:

- Papers that restricted themselves to a specific disease and/or intervention and did not generalize to eHealth within a broader context
- Papers that described the creation of a Web-based intervention and did not take the empirical evaluation in scope
- Papers on social media and support solutions that were studied separate from the Web-based intervention events
- Papers that did not describe support in functional terms (eg, praise, reassurance) but only in technical delivery terms (eg, short message service [SMS], email)
- Papers that analyzed Web-based interventions using high-level descriptive factors (eg, "interactive component," "supervision," "tailored") without going into more detail

Figure 1. Flow diagram of the study selection of part 1 of the scoping review.



Data Extraction Part 1

The entire content, including the introduction, discussion, and references, of the 22 studies was checked from the users' perspectives regarding usability and the needs they expressed. We applied grounded theory, applying the following phases:

Phase 1: We labeled the descriptions of user support needs in the way they were defined within each study.

Phase 2: We analyzed the various labels and categorized them.

Phase 3: We distilled main themes out of these categories with the aim of representing the user support content within the selected studies.

Research Question Part 2

What are main supportive features of ECAs that could potentially address user support needs?

Search Strategy Part 2: Opportunities of Virtual Coaches to Deliver Support Within Web-Based Interventions for Health or Learning

The search aimed to create a generic idea of the capabilities of ECAs for supportive purposes. The Scopus and Web of Science databases were searched with a combination of the concepts "embodied conversational agents," "Web-based intervention," and "support." For each of the concepts, multiple keywords were used (see Multimedia Appendix 1). As ECAs are often used within an e-learning context, it was decided to include

studies on intelligent tutoring systems (ITS) as well. ITS was included as a keyword within the concept of "Web-based intervention." See Textbox 2 for inclusion and exclusion criteria.

The systematic search resulted in a limited number (8) of studies. Moreover, these studies addressed a wide range of topics, from physical attributes [18], architecture [19], route planning [20], nonverbal behavior [21], virtual museum guide [22], and empathy [23], to theoretical models [24] and articulation rates [25]. None of the systematically selected studies provided a high-level picture of the capabilities of ECAs with regard to support delivery. Therefore, it was decided to expand the number of studies by means of hand search. We started the hand search in Google Scholar by both (1) checking references and (2) further searching on terms found within the 8 selected studies.

We initiated the hand search on the following basis:

1. Finding synthesizing information on ECAs within a health or pedagogical (ie, e-learning) context with a focus on the delivery of support and motivating users. We started with the information found in [23] and additionally searched for meta-studies on ECAs.
2. Finding additional (founding) studies on the computers as social actors (CASA) effect as mentioned within [18] and [23].
3. Finding additional information on relationship building [25] and measures of relationship building as briefly described in [21,25].
4. Finding additional information on theoretical models related to ECAs as touched upon in [24].

The entire search procedure resulted in including 53 studies (Figure 2).

Data Extraction Part 2

Using grounded theory, the entire content, including the introduction, discussion, and references, of the selected studies was analyzed with the aim of finding specific information on user support as carried out by ECAs. As this information was scarce, we decided to formulate three concepts that we thought were most relevant for eHealth and covered substantial information of the ECA literature that was semantically related to the notion of user support.

We formulated the following three concepts:

1. Which *multimedia* aspects of ECAs are relevant for eHealth environments?
2. What kind of *relationship* is applicable between ECAs and users?
3. How *useful* are ECAs for user adherence in eHealth?

Out of the three concepts we formulated, 8 themes coherently described a specific ECA topic.

Textbox 2. Inclusion and exclusion criteria.

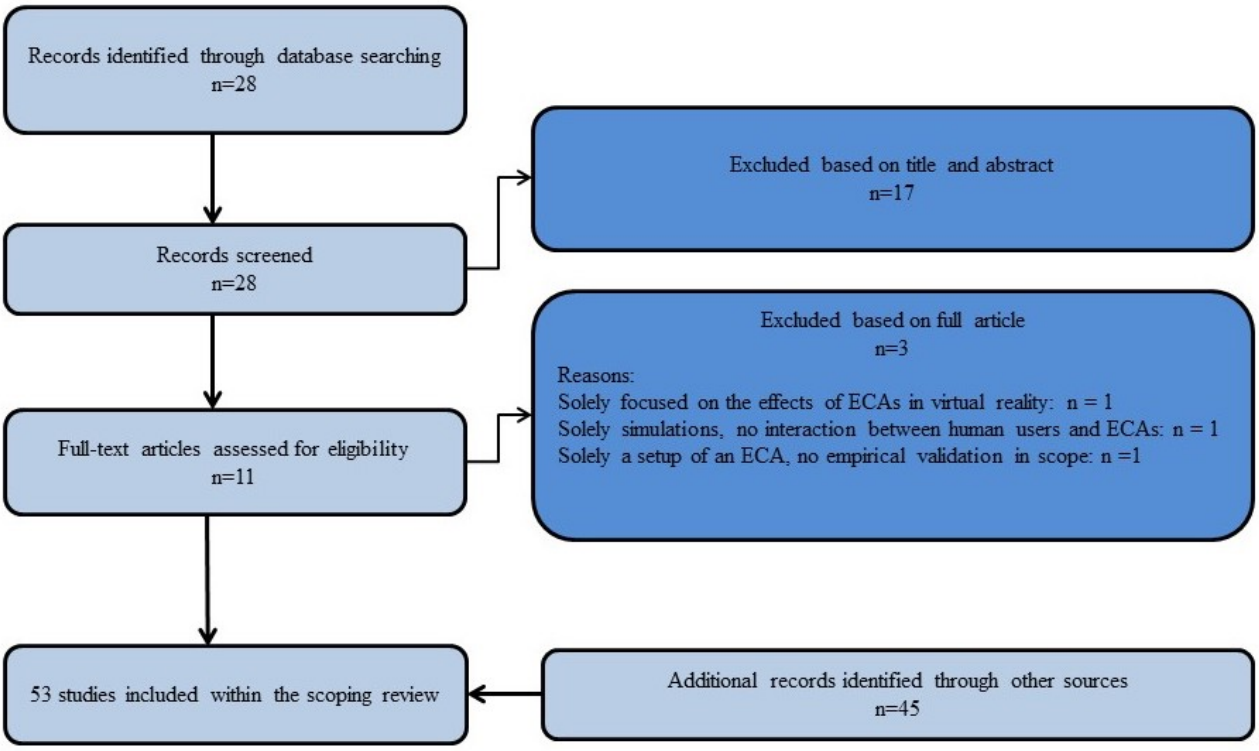
Inclusion criterion was as follows:

- Papers had to address embodied conversational agents (ECAs) interacting with users or studies on ECAs interacting with users

Exclusion criteria were as follows:

- Papers that solely focused on virtual reality
- Papers in which interaction between human users and ECAs was absent
- Papers that described the design of an ECA but did not take the empirical validation in scope

Figure 2. Flow diagram of the study selection of part 2 of the scoping review.



Results

Part 1: Results and Themes Found Within the Studies on the Need for Support in Web-Based Interventions

The 22 analyzed papers suggest that a myriad of subtle interactions between users and computers play an important role in keeping a user motivated in continuing the Web-based intervention.

We formulated 8 themes according to our data extraction procedure. We further condensed these 8 themes into 2 main need-and-support concepts that in our view summarized the subject and that would help us during further analysis.

Users expressed the need for concrete feedback on their performance. Within the literature, this need is described as the principle of closure [26]: the confirmation that an action has been successfully performed. This indicates that users of Web-based interventions could benefit from task-related interaction and support (eg, “Thank you for submitting your homework for this week. You sent it well on time.”). We call this *task-related support*.

Users expressed the need for interest and support for the issues they are dealing with. This suggests that users of Web-based interventions could benefit from emotional support that acknowledges both the user’s endeavors during the change program and the originating issue the user is dealing with. This concept was based on the literature we found earlier [12,13]. We call this *emotion-related support*.

Table 1 shows the user needs that became apparent in the included papers and that we related to the two common needs-for-support elements mentioned above.

The user needs and issues, mentioned in Table 1, are discussed in greater detail below:

Need 1: Overcome User Feelings of Isolation

The anonymity of Web-based interventions seems to play out both as a strength and weakness. Users feel encouraged to speak out but sometimes also feel isolated because of its anonymous nature. As formulated by McClay [31], “The findings showed that the desire for privacy and secrecy had a detrimental impact on participants’ help-seeking, their use of the intervention, and the support they could seek from family and friends” and, as expressed by Foster [41], “the difficulties of talking to others.”

Both task-related and emotion-related system support could potentially counteract feelings of isolation.

Need 2: Deeper Interest in the User’s Situation

Users seem to expect (and probably need) a deeper interest in their situation. Knowles et al [28] concluded the following as a shortcoming found in 7 out of 8 studies: “sensitivity to ‘Who I am’ as a patient.” Todd [32] described this as “that would deal with how to live your life as somebody who suffers from Bipolar.”

This is a case for emotion-related system support.

Table 1. User needs and issues and common user support mechanisms that can potentially fulfill these needs.

User need or issue	Support mechanism to fulfill the need	Source that describes the support mechanism
1. Overcome users' feelings of isolation	*Task-related support*[a] can conform to this need by setting and reviewing log-in goals, positively reinforcing log-in and site use and answering questions regarding the functionality of the site.	[6,27-30]
	Emotion-related support[b] can satisfy this need by establishing a supportive relationship. In case log-in goals are not met or any other sign of diminished use of the intervention appears, the system can intercede and encourage the use of the Web-based intervention.	[6,27,29,30]
2. Deeper interest in the user's situation	*Emotion-related support*[b] can fulfill this need by providing the users with the opportunity to talk about the impact of the disease on their life.	[28,31,32]
3. Interest in fundamental daily issues the user is struggling with	*Emotion-related support*[b] can meet this need by asking the users about their daily experiences and issues and by subsequently providing acknowledgment.	[28,33]
	If the user expresses a need for practical advice, the system could provide it accordingly or refer to a nurse or doctor connected to the system. This can be considered as *task-related support in a broader context*.	
4. The ability for the user to refine the communication process	*Emotion-related support*[b] provided alongside a more open interaction between user and system (eg, by means of bidirectional free text or free speech) could potentially (and so far theoretically) increase the user's feeling of contributing to his/her own change process.	[34,35]
5. The user's need for encouragement	*Emotion-related support*[b] could be delivered in terms of praising the user or delivering rewards or by other types of encouraging behavior.	[6,27,32-34,36-39]
6. Performance feedback mechanism for user responses	*Task-related support*[a] can render this need by providing corrections in case the user made factual errors (eg, homework with factual information about an illness).	[6,36,40-42]
	In addition, preferably if opted so by the user, the user's achievements can be plotted against the achievements of the user's peer group. This scenario is applicable especially for user performances that can be measured in physical terms such as step counts.	
7. Users coping with experiences of negative affect during their change process	*Emotion-related support*[b] can provide a dose of positive affect in case a phase of negative user affect that merits such a dose could be reliably distinguished.	[43]
8. Creating a setting of accountability toward the user	*Task-related support*[a] can play a positive role by objective goal setting, measuring the goals set, reminding the user of their goal set, and indicating which of these goals have (not yet) been met.	[29,30,33,38,39,44-47]

[a]Task-related support: the confirmation that a user action has been successfully performed.

[b]Emotion-related support: acknowledgement of both the user's endeavors during the change program and the originating issue the user is dealing with.

Need 3: Interest in Practical Daily Issues the User Is Struggling With

Users seem to wish for a form of deeper interest in their practical daily issues. This need is described by Knowles et al [28] as "sensitivity to 'How I Feel', recognizing the demands of depression on the user (such as emotional and motivational difficulties, and problems with concentration)," and reported on by Todd [32], "In terms of practical issues participants described wanting support understanding their legal rights, managing debt, managing pregnancy and coping with seasons and time zones."

This is a case for emotion-related system support. In case the user requests practical advice for daily issues, task-related support on these issues can also contribute.

Need 4: The Ability for the User to Refine the Communication Process

As mentioned by Donkin [34], users filling in questionnaires about how they felt said that the questionnaire did not cover their feelings. Subsequently, these users had a strong wish to contextualize their answers. McClay [31] described this as "the thought of having to fill more books in and logs."

A noninteractive tool as a questionnaire is perfectly fit for structurally gathering experimental user data. However, it may be less appreciated as it "forces" the users to answer according to its rigid structure. Emotion-related system support provided alongside a more open interaction between user and system (ie, by means of bidirectional free text or free speech) could potentially (and so far theoretically) increase the user's feeling of contributing to his or her own change process.

Need 5: The User's Need for Encouragement

As noted by Donkin et al [34] and as quoted by Mohr [30], "patients want feedback on whether they are on the 'right track' in their web-based intervention." Foster [41] described this as "Users are encouraged to set weekly SMART goals on the basis of the content of these sessions."

Encouraging users during the intervention can likely be achieved by emotion-related system support.

Need 6: Performance Feedback Mechanism for User Responses

Somewhat comparable with the statement of Donkin et al [34], Helgadóttir [42] describes that many CCBT (computerized cognitive behavioral therapy) programs would benefit from a performance feedback mechanism for user responses. This would expand the system's ability to direct the users during their change program. Gorlick [44] reported on this as "participants reported that if they were to spend time answering the questionnaires, they would prefer feedback about their responses."

Providing a direct task-related response such as "I have received your answers, thank you for your time and effort. Please allow me to comment on your answers" would immediately acknowledge the user's effort invested. By later analyzing the user responses and by providing feedback via email, a second, more profound task-related support mechanism could be implemented.

Need 7: Users Coping With Experiences of Negative Affect During Their Change Process

As formulated by Todd [32], "Participants struggled to understand why they feel this way and desired to know why lack of motivation happens and how to overcome it." Kraft et al [43] suggest that individuals should be assisted in coping with experiences of negative affect during their change process. They make a claim that many change program users struggle with the tension between their aspirations and their actual status and behavior. During this struggle, the client's internal process of self-regulation is activated to alleviate the tension. Too much burden on the self-regulation process leads to ego depletion [43], a status of a low level of mental energy. This status often results in increased relapse vulnerability and potentially therapy nonadherence. As a way to reverse this ego-depletion process, Kraft et al [43] recommend a dose of positive affect, next to a period of rest for recovery. Emotion-related system support could provide such a dose of positive affect. The challenge would be to determine the moment that ego depletion is close.

Need 8: Creating a Setting of Accountability Toward the User

As described by Bradbury [33], "Some participants' accounts of coaching suggested that they experienced accountability to the coach, which made them more committed and motivated." Mohr et al [39] also stress the importance of creating a setting of accountability toward the user, "the implicit or explicit expectation that an individual may be called upon to justify his or her actions or inactions."

For such a setting, certain preconditions are necessary, such as participants who understand and agree with the benefits of their expected future behavior. Other preconditions are concrete goal setting and performance monitoring. Task-related machine support can play a positive role by reminding the users of their goal set and by indicating which of these goals have (not yet) been met. Note that accountability might be harder to trigger among users who have been assigned to health interventions by their doctors and who did not primarily opt to participate by themselves.

Part 2: Results and Themes Found Within the Studies on ECAs With Motivational Capabilities

Table 2 shows the results and the themes that were found in the selected studies. Out of 8 themes found, 7 themes originate from the literature found during the systematic search and were bolstered during the hand search. Theme 8, methodological issues, was uniquely based on hand search information.

Theme 1: Computers as Social Actors

A large body of studies on ECAs refer to the CASA effect [49,50,51] as a cornerstone for studying human-computer interactions and especially human-ECA interactions. The CASA effect demonstrates that humans treat media—in some respect—in the same way as they treat fellow humans. Various manifestations of this effect have been described as follows:

- Computers that display flattery texts toward their users are preferred by their users compared with computers that do not display such texts.
- Computers that textually praise other computers are better liked than computers that praise themselves, and computers that *criticize* other computers are disliked compared with computers that criticize themselves.
- Users who are partnered with a computer on basis of a color (eg, the blue team) will have a more positive opinion about the computer and cooperate more with it than users who have to partner with a computer of the opposite, differently colored team.

As an explanation of the CASA effect, it has been proposed that humans have a strong innate tendency to make social connections with other humans and other living creatures such as pets. This human tendency becomes real when objects such as personal computers (PCs) demonstrate activities that could be socially interpreted by their users [51]. Although PCs can act socially, human users are logically aware of their nonsocial and nonliving status. This seems a paradox: why would a human user socially respond to a PC while at the same time realizing that a PC does not warrant it? Nass and Moon [49] refer to "mindless" (automatic, largely unaware) human behavior that the machine can trigger. This mindless behavior will be displayed as long as it remains socially acceptable. This phenomenon is also associated with the notion of "suspension of disbelief," meaning that up to a certain point humans are willing to apply social rules to nonhuman yet communicative objects, irrespective of their nonliving status.

Table 2. Themes on supportive embodied conversational agents.

Theme	Explanation	Sources
1. Computers as social actors	Humans treat media in the same way as they treat other humans.	Systematic search: [23] Hand search: [48-51] (concept=relationship)
2. Open dialogue between user and computer	Embodied conversational agents (ECAs) have the ability to have an open verbal dialogue with users.	Systematic search: [22] Hand search: [52-54] (concept=multimedia)
3. Visible conversational partner	Interaction with a "talking face" leads to more trust and believability.	Systematic search: [18,20,23,24] Hand search: [55-65] (concept=multimedia)
4. Human-ECA relationship	Interactions with an agent can lead to a relationship, which is important to keep users engaged over time.	Systematic search: [25] Hand search: [17,66-72] (concept=relationship)
5. Measures of the human-ECA relationship	Human-ECA relationship quality can be measured.	Systematic search: [17,21,68,73] (concept=relationship)
6. Responsive verbal and nonverbal communication	Computers should have the ability to notice and respond to verbally and nonverbally expressed emotions from their user to create a more natural interaction.	Systematic search: [23] Hand search: [63,74-81] (concept=relationship)
7. Impact of ECAs on user motivation	There is evidence that ECAs can motivate users, which is highly dependent on ECA implementation, context, task, etc.	Systematic search: [19] Hand search: [57,74,82-84] (concept=useful for eHealth[a])
8. Methodological issues within ECA research	Most experiments into ECAs face similar methodological issues, which have to be taken into account when interpreting the research.	Hand search: [85-89] (concept=useful for eHealth[a])

[a]eHealth: electronic health.

Theme 2: Open Dialogue Between User and Computer

The theme that follows is the ability of computers and ECAs to have an open verbal (textual or speech) dialogue with users. Within regular, day-to-day human-computer interaction events, a user who interacts with his or her information technology system will typically activate predefined menu options such as the "save as" option. Subsequently, the computer will respond to the request by presenting a pop-up window, which will enable the user to type in the file name of the document. In such a closed dialogue scenario, the interactions between the user and the software traditionally have a task-specific character (ie, serve to reach the specific goal of saving a document), have a short duration, and are typically initiated by the user (and not by the computer). In contrast, ECAs enable more open-ended and more relationship-oriented interactions. Interactions between ECAs and users can span multiple question-and-answer pairs and can therefore be interpreted as a dialogue.

The ELIZA (software created by Joseph Weizenbaum at the MIT Computer Science and Artificial Intelligence Laboratory Cambridge, MA, USA) study [54] described an early version of a textual psychotherapist that gave "canned" responses to user questions as a result of quickly processing the input text provided and create a response out of it without realizing what the user had said (eg, a question such as "Eliza, I feel miserable today" and an answer "How often do you experience feelings of being miserable?").

Later studies create richer dialogue contexts to explore the capabilities of computers interacting with humans. One of the examples is a study that has shown that a robot taking the role of a museum guide who uses, for example, empathy and humor in his conversation style led to a more positive attitude toward the robot than the same robot without this enhanced conversation style [22]. A second study showed that an ECA with high-dialogue capabilities reached more accurate answers when interviewing a subject than an agent with less dialogue capabilities [52]. A third study [53] aimed to explore where open-dialogue options between users and ECAs would lead to. The authors report that when learners are given opportunities to guide an open conversation, they especially ask off-topic questions. For example, learners often want to know about the agents' operating systems, design, purpose, and capabilities. Such conversations seem to serve the "testing" of agents' abilities during which learners are attempting to discover the boundaries, limits, and capabilities of agents through "game-like" inquiry.

Table 3. Main theories and effects of visible embodied conversational agents.

Embodied conversational agent (ECA) theory	Explanation	Source
Theory of social inhibition/facilitation	When in the presence of others, people perform learned tasks better and novel tasks worse. Empirical results have demonstrated that this principle also applies for the presence of ECAs.	[65]
Social agency theory	By adding a visible ECA as a screen tutor, the social interaction schema is primed, which will cause the learner to try to understand and deeply process the computer-delivered instructions.	[61]
Social modeling/social learning theory	Humans derive their knowledge, attitudes, behavior, and goals by observing and imitating the surrounding social agents.	[18,24]
Situational dependency	Pedagogical agents are helpful when there is a need to increase companionship and decrease complexity.	[57]
Social exchange theory	People prefer equitable relationships in which the contribution of rewards and costs are roughly equal. This equity principle also applies to human-computer relationships.	[58]
Persona effect	The presence of a lifelike character in an interactive learning environment—even one that is not expressive—can have a strong positive effect on a student's perception of his or her learning experience.	[59]
Image principle	The image of an ECA is not a key factor for learning; instead, the level of animation of the ECA is the key factor for learning.	[61]

Theme 3: Visible Conversational Partner

The next theme is the visibility of the conversational computer depicted as a (either static or animated) human face. According to Lisetti [60], the human face has a special status in human-to-human communication as it has often been identified as the most important channel for conducting trust and believability. As Lisetti states, the face as a communication channel has a higher status than bodily regions such as posture and gesture [56]. Multiple studies have supported this notion by demonstrating that users preferred to interact with a "talking face" instead of a text-only interface [64], an anthropomorphic agent together with a human voice has led to greater agent credibility [55], and visible agents have led to greater positive motivational outcomes [63] and task performance [65].

Besides empirical research, there are multiple theories that support this notion. The theories that were mentioned in the included sources are listed and explained in Table 3.

Despite these positive experimental results and theoretical support for a visible, human-like PC, the visibility subject is somewhat controversial. Strong claims against the human face are provided by Norman [26] by his statement that a human face triggers false mental models and thus creates wrong user expectations. Other critique is provided by Rajan et al [62] who demonstrated that it is first and foremost the voice (and not the visibility of the ECA) that is responsible for positive learning effects.

Theme 4: Human-ECA Relationship

The fourth theme is the concept that regular human-computer interaction events result in a relationship. Routine interactions between a user and his or her computer should be regarded as contributions to this human-computer relationship, as is argued by Bickmore et al [17]. Although this relationship may be implicit, it has an impact on the user. The relationship plays a role even in case no relationship skills (eg, empathy, humor) have been designed and built into the machine.

The question arises whether an ECA with a relationship-focused design could behave and be perceived as a competent social actor. This quality of the ECA as a conversational partner is impacted by the following:

- *Interaction duration*. As described by Krämer et al [71], getting people engaged with ECAs is easy, but keeping them engaged over time is much more challenging. Bickmore et al [17] (on physical activity) and Creed et al [67] (on fruit consumption) conducted emotional virtual coach studies that spanned more than 28 days. They both found that deploying the emotional ECA did not result in user behavior changes but that users in general preferred to interact with the emotional virtual coaches.
- *Natural versus forced interaction*. Gulz [69] suggests that most ECA studies force the human-computer relationship too much. Users have no other option than to interact with the ECAs they are confronted with.
- *User personality*. Von der Pütten et al [72] make clear that it depends on the personality of the user how the human-computer relationship will develop. They demonstrated that five user personality factors were better predictors for the evaluation outcome of ECAs than the actual behavior of the ECA.

Theme 5: Measures of the Human-ECA Relationship

The literature found mentions two regular measures with regard to the human-ECA relationship.

Measure 1: Working Alliance

Working alliance is a construct that originates from the psychotherapy literature and has been described as "the trust and belief that the helper and patient have in each other as team-member in achieving a desired outcome" [73]. Bickmore et al [66] applied the working alliance inventory in their 30-day longitudinal study with an ECA acting as an exercise coach. Participants who interacted with an ECA that was relational behavior–enabled (empathy, social chat, form of address, etc) scored the ECA significantly higher on the working alliance

inventory compared with participants who interacted with the same ECA with the relational behaviors disabled.

Measure 2: Rapport

A second important human-computer relationship measure is rapport. Rapport has been described as "the establishment of a positive relationship among interaction partners by rapidly detecting and responding to each other's nonverbal behavior" [68]. Measurement of rapport has been conducted by Gratch et al [68] in their evaluative ECA study. Their results showed that the experience of rapport was of a comparable level compared with a face-to-face (ie, human interlocutor) condition.

Theme 6: Responsive Verbal and Nonverbal Communication

Within human-to-human communication, the exchange of nonverbal information plays a key role. Social psychologists assert that more than 65% of the information exchanged during a person-to-person conversation is conveyed through the nonverbal band [75,81]. The nonverbal channel is said to be especially important to communicate socioemotional information. Socioemotional content [76] is vital for building trust and productive human relationships that go beyond the purely factual and task-oriented communication. D'Mello et al [76] describe the mutual impact of user and (synthetic) computer emotions as an affective loop, which is pictured as follows:

- The user first expresses his or her need and accompanying emotion through verbal and physical interaction with the machine, for example, through detectable gestures, usage of the keyboard, or spoken language.
- Then, the system generates an affective reply, through words, speech, and animation with the intention to respond to the user's need.
- This response affects users in such a way that they become more involved in their further interaction with the computer.

Others, such as Doirado [78], use the term "belief, desire, and interest" (BDI) in relation to a system that is (to some extent) capable of assessing the user's needs.

Concerning the importance of the affective loop and BDI, there are 2 stances:

- *Stance 1: Responsiveness of ECAs (affective loop; user BDI capable system) is a critical condition for prolonged user interaction.* Doirado et al [78] confirm the importance of the affective loop mechanism and state that an ECA that lacks the capacity to assess the user's BDI and to conform to the user's needs by adapting its behavior (a nonresponsive ECA) will break the user's suspension of disbelief.
- *Stance 2: Autonomy of ECAs (no affective loop; system is unaware of the user's BDI) is a sufficient condition for prolonged user interaction.* Rosenberg-Kima et al [63] deployed an autonomous (ie, nonresponsive) ECA that introduced itself and provided a 20-min narrative about 4 female engineers, followed by five benefits of engineering careers. The ECA was animated and its voice and lip movements were synchronized. The ECA acted autonomously; interaction between participants and ECA was purely restricted to the user clicking on the button for text topic. The results showed that the self-efficacy of the users and of their interest in the subject presented was significantly higher within the ECA + voice condition compared with the voice-only condition. In support of these results, Baylor et al [55] state that people are willing to interact with anthropomorphic agents even when their functionality is limited. As she indicates, the mere visual presence and appearance will in some contexts be the determining factor and not so much its supportive, conversational, or animation capabilities.

Theme 7: Impact of ECAs on User Motivation

Meta-studies and reviews [69,81,85,86,89] have reported on claims and evidence for positive ECAs' effects on learning, engagement, and motivation.

Schroeder et al reviewed 43 studies and concluded that pedagogical agents have a small but significant effect on learning as ultimate outcome. Within their study, Schroeder et al [81] did not make a distinction between responsive and nonresponsive ECAs. Specific research with regard to motivating users has also been conducted by deploying responsive ECAs with the task to notice user frustration and to empathically respond to it. Autonomous delivery of warmth and empathy by ECAs toward users has shown positive effects, and studies show that this effect may be larger at the time the user experiences frustration [74,86,88].

Altogether, the evidence for ECAs capable of motivating users is inconclusive. ECAs, whether they are nonresponsive or responsive, provide a positive user experience as a result of their entertainment capabilities. Responsive ECAs when specifically designed to detect user frustration and to empathically respond to it have also empirically demonstrated positive effects on user attitudes. However, these positive effects have not yet been found in ecologically valid contexts. Instead they were found within constrained contexts such as games with clear win-and-lose rules and as a result of deliberately induced user frustration.

Theme 8: Methodological Issues Within ECA Research

The inconclusiveness regarding ECA evidence as mentioned within the previous theme is claimed to be caused by methodological issues [86,89]. Methodological issues make it difficult to compare study results and to draw generic conclusions. One of those issues is the variation among ECAs. To name a few:

- Different modalities used for output: (synthesized or natural) speech or text
- Different levels of responsive emotional behavior: from textual responses projected alongside a static ECA to fine-grained ECA facial expressions intended to mirror the user's facial expressions
- Different roles: tutor, peer, interviewer, coach
- Different implementations/different computer code applied as artificial intelligence to steer the ECA

Many of these issues can be resolved by using a common, open research platform for ECAs, such as the Virtual Human platform (as provided by University of Southern California (USC) and

the Institute for Creative Technologies (ICT), Los Angeles, USA; see also [87]). Other issues can potentially be resolved by a common design framework for ECAs as proposed by Veletsianos et al with their EnALI framework [88].

Concerning the duration of the change programs, several studies (eg, [66,67]) stress that the majority of virtual coaching studies concern short time spans of tens of minutes, which makes it difficult to study the development of the human-computer relationship and to realize effects on user behavior. Both Bickmore et al and Creed et al [66,67] conducted emotional virtual coach studies that spanned more than 28 days. They both found that deploying the emotional ECA did not result in user behavior changes but that users in general preferred to interact with the emotional virtual coaches.

Altogether Dehn and van Mulken [86] summarize the situation as follows: "the simple question as to whether an animated interface improves human-computer interaction does not appear to be the appropriate question to ask. Rather, the question to ask is: what kind of animated agent used in what kind of domain influence what aspects of the user's attitudes or performance."

Discussion

Principal Findings

Research Questions

Part 1 of this scoping review addressed the following research question:

Is there a set of generic user support needs that are currently not sufficiently addressed within eHealth interventions requiring repeated use that may result in a lower user experience and therefore lower user adherence?

We found various user needs and issues related to support, which we divided into the following two main categories:

- *Task-related system support*; concrete performance-related feedback
- *Emotion-related system support*; support that has an empathic nature

It appeared that both task-related support and emotion-related support are regularly expressed user needs. Both needs therefore merit further attention in terms of research that aims to improve user adherence.

Part 2 of this scoping review addressed the following research question:

What are main supportive features of ECAs that could potentially address user support needs?

Information was scarce and a direct answer to this question could not be found. However, we were able to find relevant information on the ECA features of *multimedia*, *relationship*, and *usefulness for eHealth adherence*.

Furthermore, we made two distinctions:

- *Nonresponsive (autonomous) ECAs*. These ECAs are not designed with the intention to capture and respond to emotionally expressed user needs. These kinds of ECAs have demonstrated that they can engage users. These ECAs run the risk of annoying the user and will then become counterproductive.
- *Responsive ECAs*. These ECAs are designed with the intention to respond to user needs in real time. These ECAs have the capacity to detect and process verbal and nonverbal information uttered by humans. However, realization of these ECAs is a heavy task, requiring costly computational modeling of user BDI [77] and affective loop facilities with a high chance of failure.

Table 4 associates the needs from part 1 with the themes addressed within part 2 and indicates whether responsive or nonresponsive ECAs can address the user need.

Nonresponsive ECAs

As described within Table 4, nonresponsive ECAs can provide task-related support such as setting and reviewing log-in goals and emotion-related support by the delivery of scheduled supportive messages (need 1). Furthermore, nonresponsive ECAs are capable of motivating users by techniques such as praising (need 5), performance feedback (need 6), and setting expectation levels toward user (need 8). Altogether, nonresponsive ECAs are likely capable of helping out users with more straightforward motivational tasks.

Responsive ECAs

In contrast to nonresponsive ECAs, responsive ECAs are capable of performing more complex motivational tasks as described within the needs 3, 4, and 7. First, responsive ECAs are capable of having a dialogue with the user during which concrete daily issues the user is facing can be effectively discussed (need 3). Further research should focus on effective countermeasures for users losing interest interacting with responsive ECAs during longer-term interactions (eg, 4-10 weeks with daily contact) [65]. Second, during a dialogue with the ECA, the user can share experiences as an addition to filling in a questionnaire. This provides the user with the ability to refine the communication process (need 4). Further research should focus on the accompanying technical and conversational complexities of such a refining dialogue. Third, a responsive ECA is capable of assisting users who cope with experiences of negative affect during their change process (need 7). However, current experimental setups can only artificially create moments of frustration. Further research should focus on ECAs that detect and respond to spontaneous user emotion.

Not Addressable by Either Responsive or Nonresponsive ECAs

Dialogues between the user and ECA on deep, personal issues (need 2) are currently technically too complex to realize. Smooth interactions are a necessary condition for ECAs to become and remain a trustworthy counterpart. None of the ECAs found are capable of truly meeting this condition of smoothness. As a result of future progress within the artificial intelligence field, this may change for the better. For the moment, these dialogues should be best carried out by a human support provider.

Table 4. User needs with supportive elements, associated embodied conversational agent (ECA) features, and the needed level of responsiveness of the ECA.

User need or issue	Supportive element	Associated ECA features	Needed responsiveness
1. Overcome users' feelings of isolation	*Task-related support* can fulfill this need by setting and reviewing log-in goals . *Emotion-related support* can fulfill this need by establishing a supportive relationship,	Computers as social actors; visible conversation partner; human-computer relationship	A nonresponsive embodied conversational agent (ECA) is sufficient
2. Deeper interest in the user's situation	*Emotion-related support* can fulfill this need by providing the user with the opportunity to talk about the impact of the disease on having become a patient.	Computers as social actors; open dialogue; visible conversation partner; human-computer relationship; responsive verbal and nonverbal communication	No ECA is currently likely to be able to address this user need
3. Interest in fundamental daily issues the user is struggling with	*Emotion-related support* can fulfill this need by asking the users about their daily experiences and issues.	Computers as social actors; open dialogue; visible conversation partner; human-computer relationship; responsive verbal and nonverbal communication	A responsive ECA is necessary; further research is advised
4. The ability for the user to refine the communication process	*Emotion-related support* can be provided alongside a more open interaction between the user and the system.	Open dialogue	A responsive ECA is necessary; further research is advised
5. The user's need for encouragement	*Emotion-related support* could be delivered in terms of, for example, praising the user.	Motivational effects	A nonresponsive ECA is sufficient
6. Performance feedback mechanism for user responses	*Task-related support* can fulfill this need by reviewing the user's contributions and by providing corrections in case the user made factual errors.	Computers as social actors; visible conversation partner; human-computer relationship	A nonresponsive ECA is sufficient
7. Users coping with experiences of negative affect during their change process	*Emotion-related support* can be extended in the sense of providing a dose of positive affect at the right moment.	Responsive verbal and nonverbal communication; motivational effects	A responsive ECA is necessary; further research is advised
8. Creating a setting of accountability toward the user	*Task-related support* can play a positive role by objective goal setting.	Computers as social actors; human-computer relationship	A nonresponsive ECA is sufficient

Design Factors for Both Responsive and Nonresponsive ECAs

The ECA literature of part 2, for example [56,74], gave indications on successful design of ECAs. Some design factors have generic relevance, irrespective of deploying either a responsive or a nonresponsive ECA within an eHealth intervention. First, it is recommended [74] that the ECA communicates its *intention*, *capabilities*, and *limitations*. That is, the ECA presents itself (eg, as a coach, tutor, or peer) before the start of the intervention and behaves according to its role consistently. This way, the user will have clear expectations of the ECA's role. Second, users should have control over the presence of the ECA, especially during longer-term interactions. This will avoid user annoyance as reported by Bickmore et al [17]. Third, it is recommended that the ECA has short dialogues with the user. Systems that permit longer open-ended dialogues are playfully tested [52]. By limiting the scope and length of the dialogues, the ECA will more likely keep up its credibility.

Limitations

This review has several limitations. Due to the nature of this study as a scoping review, no quantitative analyses were done, and selection of the studies was done by interpretation of the researchers. No exclusion criteria were applied with regard to the quality of the studies to ensure broad coverage of the studied topics.

As we looked for generic user support needs in part 1, we did not take the type of mental and/or physical disorder into account. In addition, we left out factors such as user personality. The rationale was to separate the subject of user experience from the user's characteristics, but it is not certain that this separation always holds. Although we included user experience in our search string, we left out more fine-grained search terms, such as for user-centered design, to keep the search focused on the core issues. This focus on generic user needs has resulted in a broad overview of the needs and the possibilities of ECAs to address these needs, but when designing Web-based intervention for a specific target group, more research is needed to understand their specific needs for support.

Within part 2, we were aiming for on-screen solutions that could be added to the eHealth environments in *practice.* As a result, we left out studies on human-robot interaction (requiring an off-screen robot), studies with a focus on Wizard-of-Oz solutions (during which scholars steer the ECAs), and studies on virtual reality (requiring special glasses). We do not mean to imply that these technologies are not potentially interesting but only that they are less practical in the context of Web-based interventions.

Conclusions

We conclude that users of self-guided eHealth interventions can likely profit from the support of *nonresponsive* ECAs for small motivational issues. Nonresponsive ECAs that explicitly express

their supportive intention and act accordingly make self-guided eHealth environments a more user-friendly experience. This is likely to pay out in terms of adherence.

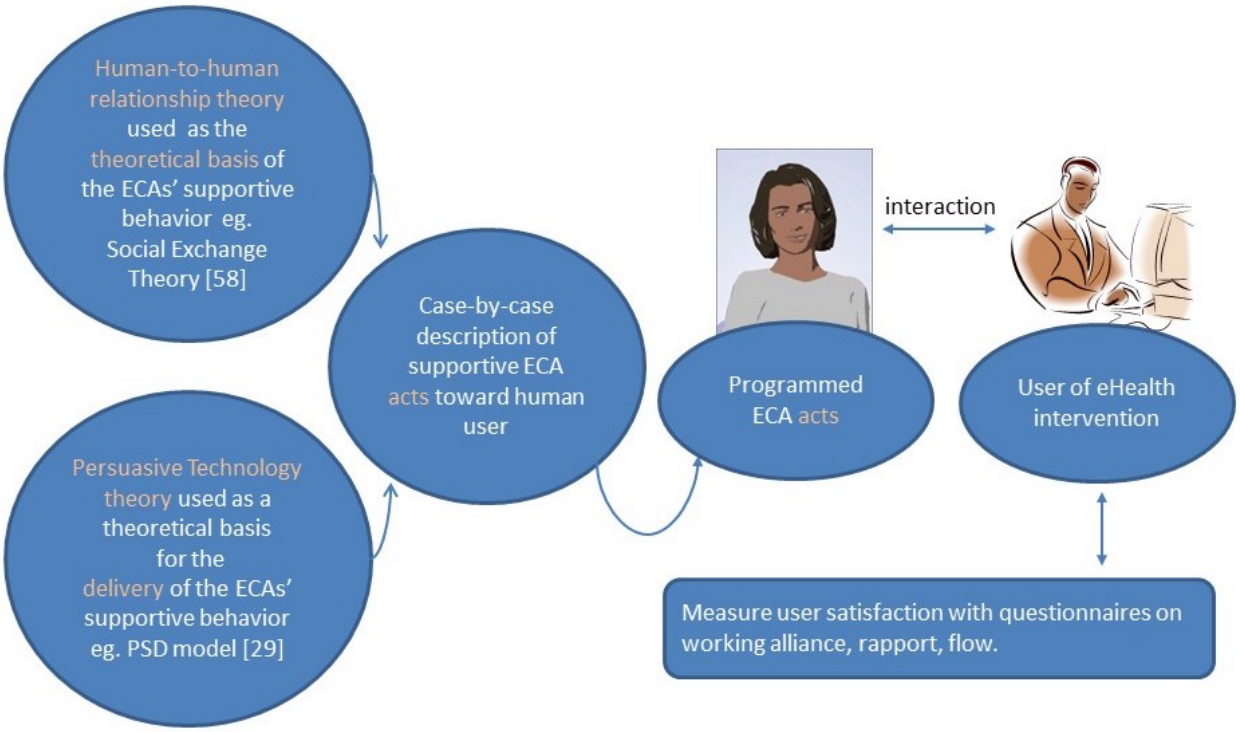
Responsive ECAs are expected to be capable of dealing with more profound motivational issues. This will require a sophisticated technological design, with sensors to capture user emotions in real time, artificial intelligence for interpretation, and speech facilities for smooth replies. The concepts of assessing a user's BDI and of deploying the affective loop to resolve user frustration are intriguing. They fit with the concept of counteracting ego depletion as addressed within part 1, theme 7.

Responsive ECAs are also relevant from other perspectives. Psychological experiments extensively make use of questionnaires to gather user data. As touched upon in part 1, need 4, questionnaires are structured yet limited communication tools by design. The ECA's sensors that deliver real-time signals on the user's BDI during experiments can provide an additional source of user information for analysis. As an alternative to the sensor and artificial intelligence technology working in real time, logs on intervention usage (eg, number of log-ins, time in between log-ins) could predict lower user motivation.

To successfully *motivate* the user, the ECA should make use of relationship theories. The social exchange theory suggested by Krämer et al (see Table 3) provides a promising example. Application of this theory to eHealth suggests that humans prefer *equitable* human-computer relationships in which the contribution of rewards and costs are roughly equal. For an eHealth intervention, a dose of positive encouragement may serve as an effective counterbalance to the user's effort invested. Put differently, where the eHealth intervention not only *demands* but also *provides* support, the human-computer relationship may be more equitable. Such an equipollent relationship will hypothetically last longer, as is exemplified among humans. Reversely, research on human-to-human relationship theories can profit from research on responsive ECAs, that is, an ECA that bases its acts on a human-to-human relationship theory makes this theory potentially verifiable. Testing the effects of the ECA on a user can contribute to a deeper understanding of the relationship theory.

Finally, we would like to propose a research framework. Following the advice of Dehn and van Mulken [86] to be specific about ECAs, our framework describes a *supportive ECA as an adjunct to an eHealth or ITS solution*. As Figure 3 depicts, the ECA is theory based, both from a *relationship* perspective and from a *persuasive technology* perspective. These theories lead to *supportive* ECA *acts*, as realized during the *programming phase*. When running an experiment on ECA-user interaction, the user's level of satisfaction is measured. This is done *postexperimentally* by means of questionnaires. Additionally, *intermediary* user signals (not depicted) can be captured by sensors and analyzed. This information can be fed into the responsive ECA to make its behavior adaptive. It can also be used to cross-check the questionnaires. As expressed by Michie et al [90], eHealth interventions should be theory based. We would like to add, so should ECAs that support their users.

Figure 3. Proposal for a theory-based framework for supportive electronic health (eHealth) embodied conversational agents (ECAs).



Conflicts of Interest

None declared.

Multimedia Appendix 1

Keywords search part 1 and 2.

[PDF File (Adobe PDF File), 28KB - jmir_v19i11e383_app1.pdf]

References

1. Andrews G, Cuijpers P, Craske M, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2013 Oct 13;5(10):e3196. [Medline: 20967242]
2. Cuijpers P, Donker T, Johansson R, Mohr DC, van Straten A, Andersson G. Self-guided psychological treatment for depressive symptoms: a meta-analysis. PLoS One 2011 Jun 21;6(6):1274. [Medline: 21712998]
3. Richardson T, Richards D. Computer-based psychological interventions for depression treatment: a systematic review and meta-analysis. Clin Psychol Rev 2012;32(4):329-342. [Medline: 22466510]
4. Barak A, Klein B, Proudfoot JG. Defining Internet-supported therapeutic interventions. Ann Behav Med 2009;38(1):4-17. [Medline: 19787305]
5. Andersson G, Carlbring P, Berger T, Almlöv J, Cuijpers P. What makes Internet therapy work? Cogn Behav Ther 2009 Feb 03;38 Suppl 1:55-60. [doi: 10.1080/16506070902916400] [Medline: 19675956]
6. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152. [Medline: 23151820]
7. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [Medline: 21164337]
8. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb 14;13(1):e17. [Medline: 21320854]
9. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16. [Medline: 19632970]
10. Talbot F. Client contact in self-help therapy for anxiety and depression: necessary but can take a variety of forms beside therapist contact. Behav Change 2012 Jul 01;29(2):63-76. [doi: 10.1017/bec.2012.4]
11. Newman M, Szkodny L, Llera S, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011;31(1):89-103. [Medline: 21130939]
12. Cavanagh K, Millings A. (Inter) personal computing: the role of the therapeutic relationship in e-mental health. J Contemp Psychother 2013 Jul 17;43(4):197-206. [doi: 10.1007/s10879-013-9242-z]
13. Kreijns K, Kirschner PA, Jochems W. Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. Comput Human Behav 2003 Jan 16;19(3):335-353. [doi: 10.1016/S0747-5632(02)00057-2]
14. Bickmore T. Relational agents for chronic disease self management. In: Hayes BM, Aspray W, editors. Health Informatics: A Patient-Centered Approach to Diabetes. Cambridge, MA: MIT Press; 2010:181-204.
15. Pham MT, Rajic A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014;5(4):371-385. [Medline: 26052958]
16. Krämer NC, Bente G. Personalizing e-Learning. The social effects of pedagogical agents. Educ Psychol Rev 2010;22(1):71-87. [doi: 10.1007/s10648-010-9123-x]
17. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
18. Jordine K, Wilson DM, Sakpal R. What is age's affect in collaborative learning environments? In: Stephanidis C, Antona M, editors. Universal Access in Human-Computer Interaction. User and Context Diversity. UAHCI 2013. Lecture Notes in Computer Science, vol 8010. Berlin, Heidelberg: Springer; 2013.
19. Ieronutti L, Chittaro L. Employing virtual humans for education and training in X3D/VRML worlds. Comput Educ 2005;49(1):93-109. [doi: 10.1016/j.compedu.06]
20. Hofmann H, Tobisch V, Ehrlich U, Berton A. Evaluation of speech-based HMI concepts for information exchange tasks: a driving simulator study. Comput Speech Lang 2015 Sep 01;33(1):109-135. [doi: 10.1016/j.csl.01]
21. Novick D, Gris I. Building rapport between human and ECA: a pilot study. In: Kurosu M, editor. Human-Computer Interaction. Advanced Interaction Modalities and Techniques. HCI 2014. Lecture Notes in Computer Science, vol 8511. Cham: Springer; 2014.
22. Bickmore TW, Vardoulakis LM, Schulman D. Tinker: a relational agent museum guide. Auton Agent Multi-Agent Syst 2013;27(2):254-276. [doi: 10.1007/s10458-012-9216-7]

23. Amini R, Lisetti C, Yasavur U, Rishe N. On-demand virtual health counselor for delivering behavior-change health interventions. In: Proceedings of the Healthcare Informatics (ICHI).: IEEE; 2013 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); September 9, 2013; Philadelphia, PA, USA p. 46-55.
24. Apostol S, Soica O, Manasia L, Stefan C. Virtual pedagogical agents in the context of virtual learning environments: framework and theoretical models. Elearn Softw Educ 2013;2:531-536.
25. Schulman D, Bickmore T. Modeling behavioral manifestations of coordination and rapport over multiple conversations. In: Proceedings of the International Conference on Intelligent Virtual Agents. 2010 Sep 20 Presented at: IVA: International Conference on Intelligent Virtual Agents; September 20-22, 2010; Philadelphia, PA, USA.
26. Norman D. Emotional Design Why We Love (or Hate) Everyday Things. New York: Basic Books; 2004:1-257.
27. Kelders SM, Kok RN, Van Gemert-Pijnen JE. Technology and adherence in web-based interventions for weight control: a systematic review. 2011 Presented at: Proceedings of the 6th International Conference on Persuasive Technology. Persuasive Technology and Design: Enhancing Sustainability and Health; June 2-5, 2011; Columbus, Ohio, USA p. 1-8.
28. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014 Jan 17;9(1):1-10. [doi: 10.1371/journal.pone.008]
29. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011;13(3):e46.
30. Mohr D, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086. [Medline: 23990896]
31. McClay CA, Waters L, McHale C, Schmidt U, Williams C. Online cognitive behavioral therapy for bulimic type disorders, delivered in the community by a nonclinician: qualitative study. J Med Internet Res 2013;15(3):e46. [Medline: 23502689]
32. Todd NJ, Jones S, Lobban F. What do service users with bipolar disorder want from a web-based self-management intervention?: a qualitative focus group study. Clin Psychol Psychother 2013;20(6):531-543. [doi: 10.1002/cpp.1804]
33. Bradbury K, Dennison L, Little P, Yardley L. Using mixed methods to develop and evaluate an online weight management intervention. Brit J Health Psych 2015;20(1):45-55. [doi: 10.1111/bjhp.1]
34. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91. [doi: 10.2196/jmir.2100]
35. Nicholas J, Proudfoot J, Parker G, Gillis I, Burckhardt R, Manicavasagar V, et al. The ins and outs of an online bipolar education program: a study of program attrition. J Med Internet Res 2010;12(5):e57. [Medline: 21169169]
36. Cowpertwait L, Clarke D. Effectiveness of web-based psychological interventions for depression: a meta-analysis. Int J Ment Health 2013;11(2):247-268. [doi: 10.1007/s11469-012-9416-z]
37. Jalil S, Myers T, Atkinson I. A meta-synthesis of behavioral outcomes from telemedicine clinical trials for type 2 diabetes and the Clinical User-Experience Evaluation (CUE). J Med Syst 2015;39(3):28. [Medline: 25677954]
38. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013;15(2):e37. [Medline: 23425685]
39. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30. [Medline: 21393123]
40. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251. [Medline: 24332469]
41. Foster C, Calman L, Grimmett C, Breckons M, Cotterell P, Yardley L, et al. Managing fatigue after cancer treatment: development of RESTORE, a web-based resource to support self-management. Psychooncology 2015 Aug;24(8):940-949. [doi: 10.1002/pon.3747] [Medline: 25648410]
42. Helgadottir FD, Menzies RG, Onslow M, Packman A, O'Brian SO. Online CBT I: bridging the gap between Eliza and modern online CBT treatment packages. Behav Change 2009;26(4):245-253.
43. Kraft P, Drozd F, Olsen E. Digital therapy: addressing willpower as part of the cognitive-affective processing system in the service of habit change. 2008 Presented at: Proceedings of the 3rd International Conference on Persuasive Technology; June 4-6, 2008; Oulu, Finland p. 177-188.
44. Gorlick A, Bantum EO, Owen JE. Internet-based interventions for cancer-related distress: exploring the experiences of those whose needs are not met. Psychooncology 2013 Nov 17;23(4):452-458. [doi: 10.1002/pon.3443]
45. Hamel LM, Robbins LB. Computer- and web-based interventions to promote healthy eating among children and adolescents: a systematic review. J Adv Nurs 2012;69(1):16-30. [doi: 10.1111/j.1365-2648.06086.x] [Medline: 22757605]
46. Ramadas A, Quek KF, Chan CK, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011;80(6):389-405. [Medline: 21481632]
47. Sinclair C. Effectiveness and user acceptance of online chronic disease management interventions in rural and Remote setting: systematic review and narrative synthesis. Clin Med Insights Ther 2015;7:43-52. [doi: 10.4137/CMT.S18553]
48. Mori M, MacDorman KF, Kageki N. The uncanny valley [from the field]. IEEE Robot Autom Mag 2012;19(2):98-100. [doi: 10.1109/MRA.2012.2192811]

49. Nass C, Moon Y. Machines and mindlessness: social responses to computers. J Soc Isssues 2000 Jan;56(1):81-103. [doi: 10.1111/0022-4537.00153]
50. Nass C, Brave S. Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. Cambridge, MA: MIT press; 2005.
51. Reeves B, Nass C. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places (CSLI Lecture Notes S). Cambridge, UK: CSLI Publications and Cambridge University Press; 1996.
52. Conrad FG, Schober MF, Jans M, Orlowski RA, Nielsen D, Levenstein R. Comprehension and engagement in survey interviews with virtual agents. Front Psychol 2015;6:1578. [doi: 10.3389/fpsyg.2015.01578]
53. Veletsianos G, Russell GS. What do learners and pedagogical agents discuss when given opportunities for open-ended dialogue? J Educ Comput Res 2013;48(3):381-401. [doi: 10.2190/EC.48.3.e]
54. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966;9(1):36-45. [doi: 10.1145/365153.365168]
55. Baylor AL, Ryu J, Shen E. The effects of pedagogical agent voice and animation on learning, motivation and perceived persona. : June 23-28, 2003; 2003 Presented at: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications; 2003; Honolulu, Hawaii, USA p. 452-458.
56. Cowell AJ, Stanney KM. Embodiment and interaction guidelines for designing credible, trustworthy embodied conversational agents. In: Rist T, Aylett RS, Ballin D, Rickel J, editors. Intelligent Virtual Agents. IVA 2003. Lecture Notes in Computer Science, vol 2792. Berlin, Heidelberg: Springer; 2003:301-309.
57. Kim C, Baylor AL. A virtual change agent: motivating pre-service teachers to integrate technology in their future classrooms. J Educ Technol Soc 2008;11(2):309-321 [FREE Full text]
58. Krämer NC, von der Pütten A, Eimler S. Human-agent and human-robot interaction theory: similarities to and differences from human-human interaction. In: Zacarias M, de Oliveira JV, editors. Human-Computer Interaction: The Agency Perspective. Berlin, Heidelberg: Springer; 2012:215-240.
59. Lester JC, Converse SA, Kahler S, Barlow S, Stone B, Bhogal R. The persona effect: affective impact of animated pedagogical agents. 1997 Presented at: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems; March 22-27, 1997; Atlanta, Georgia, USA p. 359-366.
60. Lisetti C, Amini R, Yasavur U. Now all together: overview of virtual health assistants emulating face-to-face health interview experience. Künstl Intell 2015;29(2):161-172. [doi: 10.1007/s13218-015-0357-0]
61. Mayer RE. Cognitive theory of multimedia learning. In: The Cambridge Handbook of Multimedia Learning. Cambridge, UK: Cambridge University Press; 2014:43-71.
62. Rajan S, Craig SD, Gholson B, Person NK, Graesser AC, The Tutoring Research Group. AutoTutor: Incorporating back-channel feedback and other human-like conversational behaviors into an intelligent tutoring system. Int J Speech Technol 2001;4(2):117-126. [doi: 10.1023/A:1017319110294]
63. Rosenberg-Kima RB, Baylor AL, Plant EA, Doerr CE. The importance of interface agent visual presence: voice alone is less effective in impacting young women's attitudes toward engineering. 2007 Presented at: International Conference on Persuasive Technology; April 26-27, 2007; Palo Alto, CA, USA p. 214-222.
64. Sproull L, Subramani M, Kiesler S, Walker JH, Waters K. When the interface is a face. Human Comput Interaction 1996 Jun 1;11(2):97-124. [doi: 10.1207/s15327051hci1102_1]
65. Zanbaka C, Ulinski A, Goolkasian P, Hodges LF. Effects of virtual human presence on task performance. 2004 Presented at: Proceedings of the International Conference on Artificial Reality and Telexistence (ICAT); 2004; Coex, Korea p. 174-181.
66. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput Hum Interact 2005;12(2):293-327. [doi: 10.1145/1067860.1067867]
67. Creed C, Beale R, Cowan B. The impact of an embodied agent's emotional expressions over multiple interactions. Interact Comput 2015 Mar;27(2):172-188. [doi: 10.1093/iwc/iwt]
68. Gratch J, Wang N, Gerten J, Fast E, Duffy R. Creating rapport with virtual agents. 2007 Sep Presented at: IVA 2007: Intelligent Virtual Agents; September 17-19, 2007; Paris, France p. 125-138.
69. Gulz A. Benefits of virtual characters in computer based learning environments: claims and evidence. Int J Artif Intell Educ 2004;14(3,4):313-334.
70. Kang SH, Gratch J, Wang N, Watt JH. Does the contingency of agents' nonverbal feedback affect users' social anxiety? 2008 May Presented at: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1; 2008; May 12-16, 2008; Estoril, Portugal p. 120-127.
71. Krämer N, Eimler S, Rosenthal-von der Pütten AM, Payr S. Theory of companions: what can theoretical models contribute to applications and understanding of human-robot interaction? Appl Artif Intell 2011;25(6):474-502. [doi: 10.1080/08839514.2011.587153]
72. von der Pütten A, Kramer N, Gratch J. How our personality shapes our interactions with virtual characters-implications for research and development. 2010 Presented at: Intelligent Virtual Agents; September 20-22, 2010; Philadelphia, PA, USA p. 2010-2221.
73. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Couns Psychol 1989;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]

74. Baylor AL, Rosenberg-Kima RB, Plant EA. Interface agents as social modelsthe impact of appearance on females' attitude toward engineering. 2006 Presented at: Proceedings of the CHI'06 Extended Abstracts on Human Factors in Computing Systems; April 22-27, 2006; Montréal, Québec, Canada p. 526-531.
75. Berry D, Butler LT, de Rosis F. Evaluating a realistic agent in an advice-giving task. Int J Hum Comput Stud 2005 Sep;63(3):304-327. [doi: 10.1016/j.ijhcs.2005.03.006]
76. D'Mello S, Picard RW, Graesser A. Towards an affect-sensitive autotutor. IEEE Intell Syst 2007;22(4):53-61. [doi: 10.1109/MIS.2007.79]
77. D'Mello S, Olney A, Williams C, Hays P. Gaze tutor: a gaze-reactive intelligent tutoring system. Int J Hum Comput Stud 2012;70(5):004. [doi: 10.1016/j.ijhcs.2012.01.004]
78. Doirado E, Martinho C. I mean it!: detecting user intentions to create believable behaviour for virtual agents in games. 2010 Presented at: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems; May 10-14, 2010; Toronto, Canada p. 83-90.
79. Lisetti C, Amini R, Yasavur U, Rishe N. I can help you change! An empathic virtual agent delivers behavior change health interventions. ACM Trans Manag Inf Syst 2013;4(4):19. [doi: 10.1145/2544103]
80. Picard RW. Affective Computing. Cambridge, MA: MIT Press; 1997.
81. Schroeder NL, Adesope OO, Gilbert RB. How effective are pedagogical agents for learning? A meta-analytic review. J Educ Comput Res 2013;49(1):1-39. [doi: 10.2190/EC.49.1.a]
82. Brave S, Nass C, Hutchinson K. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int J Human Comput Stud 2005 Feb;62(2):161-178. [doi: 10.1016/j.ijhcs.2004.11.002]
83. Klein J, Moon Y, Picard RW. This computer responds to user frustration: theory, design, and results. Interact Comput 2002 Feb;14(2):119-140. [doi: 10.1145/632716.632866]
84. Moreno R. Software agents in multimedia: an experimental study of their contributions to students? learning. 2001 Presented at: Proceedings of Human-Computer Interaction International; August 5-10, 2001; New Orleans, USA p. 275-277.
85. Beale R, Creed C. Affective interaction: how emotional agents affect users. Int J Hum Comput Stud 2009;67(9):755-776. [doi: 10.1016/j.ijhcs.2009.05.001]
86. Dehn DM, van Mulken S. The impact of animated interface agents: a review of empirical research. Int J Hum Comput Stud 2000;52(1):1-22. [doi: 10.1006/ijhc.1999.0325]
87. Hartholt A, Traum D, Marsella SC, Shapiro A, Stratou G, Leuski A, et al. All together now. 2013 Presented at: Proceedings of the International Workshop on Intelligent Virtual Agents; August 29-31, 2013; Edinburgh, UK.
88. Veletsianos G, Miller C, Doering A. EnALI: A research and design framework for virtual characters and pedagogical agents. J Educ Comput Res 2009;41(2):171-194. [doi: 10.2190/EC.41.2.c]
89. Veletsianos G, Russell GS. Pedagogical agents. In: Spector JM, Merrill MD, Elen J, Bishop MJ, editors. Handbook of Research on Educational Communications and Technology. New York: Springer Nature; 2014:759-769.
90. Michie S, van Straten MM, West R. The behaviour change wheel a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42. [doi: 10.1186/1748-5908-6-42]

Abbreviations

BDI: belief, desire, and interest
ECA: embodied conversational agent
eHealth: electronic health
ITS: intelligent tutoring system

Edited by G Eysenbach; submitted 18.01.17; peer-reviewed by O Blanson Henkemans, W Keijser; comments to author 26.03.17; revised version received 24.05.17; accepted 08.09.17; published 16.11.17.

Please cite as:
Scholten MR, Kelders SM, Van Gemert-Pijnen JEWC
Self-Guided Web-Based Interventions: Scoping Review on User Needs and the Potential of Embodied Conversational Agents to Address Them
J Med Internet Res 2017;19(11):e383
URL: http://www.jmir.org/2017/11/e383/
doi:10.2196/jmir.7351
PMID:29146567

Original Paper

Reduction in Vegetable Intake Disparities With a Web-Based Nutrition Education Intervention Among Lower-Income Adults in Japan: Randomized Controlled Trial

Saki Nakamura[1,2]; Takayo Inayama[1]; Kazuhiro Harada[3]; Takashi Arao[4]

[1]Department of Health Promotion Sciences, Graduate School of Human Health Sciences, Tokyo Metropolitan University, Tokyo, Japan

[2]Research Fellow of Japan Society for the Promotion of Science, Tokyo, Japan

[3]Graduate School of Human Development and Environment, Kobe University, Hyogo, Japan

[4]Faculty of Sport Sciences, Waseda University, Saitama, Japan

Corresponding Author:
Takayo Inayama
Department of Health Promotion Sciences
Graduate School of Human Health Sciences
Tokyo Metropolitan University
Minami-Osawa 1-1
Hachioji
Tokyo, 192-0397
Japan
Phone: 81 426771111 ext 4664
Email: tinayama@tmu.ac.jp

Abstract

Background: No existing Web-based nutrition education interventions have been evaluated in light of socioeconomic status just in Japan.

Objective: The aim was to investigate the effect of a Web-based intervention program on reducing vegetable intake disparities between low- and middle-income Japanese adults.

Methods: In this randomized controlled trial, participants were assessed at three time points—baseline, postintervention (5 weeks later), and a follow-up after 3 months—from October 2015 to March 2016. We collected data via a Japanese online research service company from 8564 adults aged 30 to 59 years. Participants were stratified according to national population statistics for gender and age, and randomly selected. They were then randomly allocated into intervention (n=900) and control (n=600) groups such that both groups contained an equal number of individuals with low and middle income. The intervention program encouraged behavior change using behavioral theories and techniques tailored to their assumed stage of change. The outcome was vegetable intake servings per day (1 serving being approximately 70 g).

Results: Out of 900 participants who started, 450 were from the middle income group (of which 386 or 85.7% completed the intervention), and 450 were from the low income group (of which 371 or 82.4% completed). In the intervention group, vegetable intake increased in the low-income participants from baseline to postintervention (0.42 servings, 95% CI 0.11-0.72). A two-way analysis of variance showed that low-income participants had significant main effects of group ($\eta 2$=0.04, P=.01) and time ($\eta 2$=0.01, P<.001), and a significant interaction ($\eta 2$=0.01, P=.009). Middle-income participants also had a significant main effect of time ($\eta 2$=0.01, P=.006) and a significant interaction ($\eta 2$=0.01, P=.046).

Conclusions: This Web-based nutritional education intervention could fill the vegetable intake gap between low- and middle-income adults in Japan, and is expected to prevent noncommunicable and lifestyle-related diseases. Further intervention program improvements are necessary to maintain and increase vegetable intake for other groups.

Trial Registration: Current Controlled Trials (UMIN-ICDR): UMIN000019376; https://upload.umin.ac.jp/cgi-open-bin/icdr_e/ctr_view.cgi?recptno=R000022404 (Archived by WebCite at http://www.webcitation.org/6u9wihBZU)

(J Med Internet Res 2017;19(11):e377) doi:10.2196/jmir.8031

KEYWORDS

vegetable intake; adults; socioeconomic disadvantage; Web-based nutrition intervention; randomized controlled trial

Introduction

Background

Reducing health disparities is important for public health promotion [1]. Disparities in food intake are known to occur among socioeconomically disadvantaged people [2-4]. Appropriate vegetable intake prevents cancer [5] and obesity [6], and reduces the risk of cardiovascular disease [7-9] and other lifestyle-related diseases. Despite this, individuals with low socioeconomic status (SES) tend to have low vegetable intake [10]. Thus, promoting vegetable intake in low-SES individuals to reduce health disparities is important globally.

Nutritional and Dietary Problems in Japan

Japan has one of the highest levels of longevity in the world. However, recently, health disparities have been recognized as a social problem [11,12]. Health Japan 21 [13] recommends a vegetable intake of 350 g (5 servings) per day for adults to reduce health disparities related to lifestyle-related diseases. However, low-income people tend to consume few vegetables [10] (in the lowest income bracket: men 254 g per day, women 282 g per day). In a cross-sectional study of Japanese adults, a low percentage of individuals with lower annual income (<¥3,000,000, which was equivalent to approximately US $24,987 in October 2015) ate five servings (approximately 350 g) of vegetables daily: men 5.5% and women 10.4% [14]. Currently, practical strategies for reducing vegetable intake disparities are lacking and, therefore, are urgently needed.

A Theory Suitable for Nutrition Education

A systematic review revealed that research has utilized multiple health behavior theories in attempting to increase vegetable intake, such as stages of change [15], social cognitive theory [16], the theory of planned behavior [17], and technology-based behavior change models [18,19]. Henry et al [20] suggested the possibility of increasing vegetable intake in low-income women by using a nutrition education intervention focusing on improving self-efficacy (perceived behavioral control) [21]. Thus, the gap in vegetable intake between low- and middle-income individuals might be reduced through a multicomponent nutrition education program that focuses on self-efficacy. It is therefore necessary to evaluate the nutrition education program developed in terms of whether it produces the intended outcome in evaluating nutrition education based on a multicomponent nutrition education program, not only outcome evaluation but also process evaluation, such as perceived behavioral control.

Prior Work

There are some concerns about applying Web-based interventions to socioeconomically disadvantaged populations because they might have access only to poorer quality Internet environments. Nevertheless, Web-based interventions are generally easier to access, lower cost, and tend to be comfortable for most users. They similarly have advantages in being able to provide standardized information regardless of place or population size. These interventions have been drawing attention in recent years, with many studies confirming their efficacy in health promotion in adults [22-27]. For instance, Web-based interventions were able to increase vegetable intake in low-SES adults in rural America [28,29]. However, these studies did not examine reductions in vegetable intake disparities because they focused only on individuals with low SES.

Objectives of This Study

Our study was designed to investigate reductions in vegetable intake disparities between low- and middle-income adults. We developed a Web-based nutrition education program that incorporates multiple health behavior theories to promote vegetable intake [30]. The aim was to investigate the effects of this program on the vegetable intake and patterns of change in vegetable intake of low- and middle-income adults in Japan.

Methods

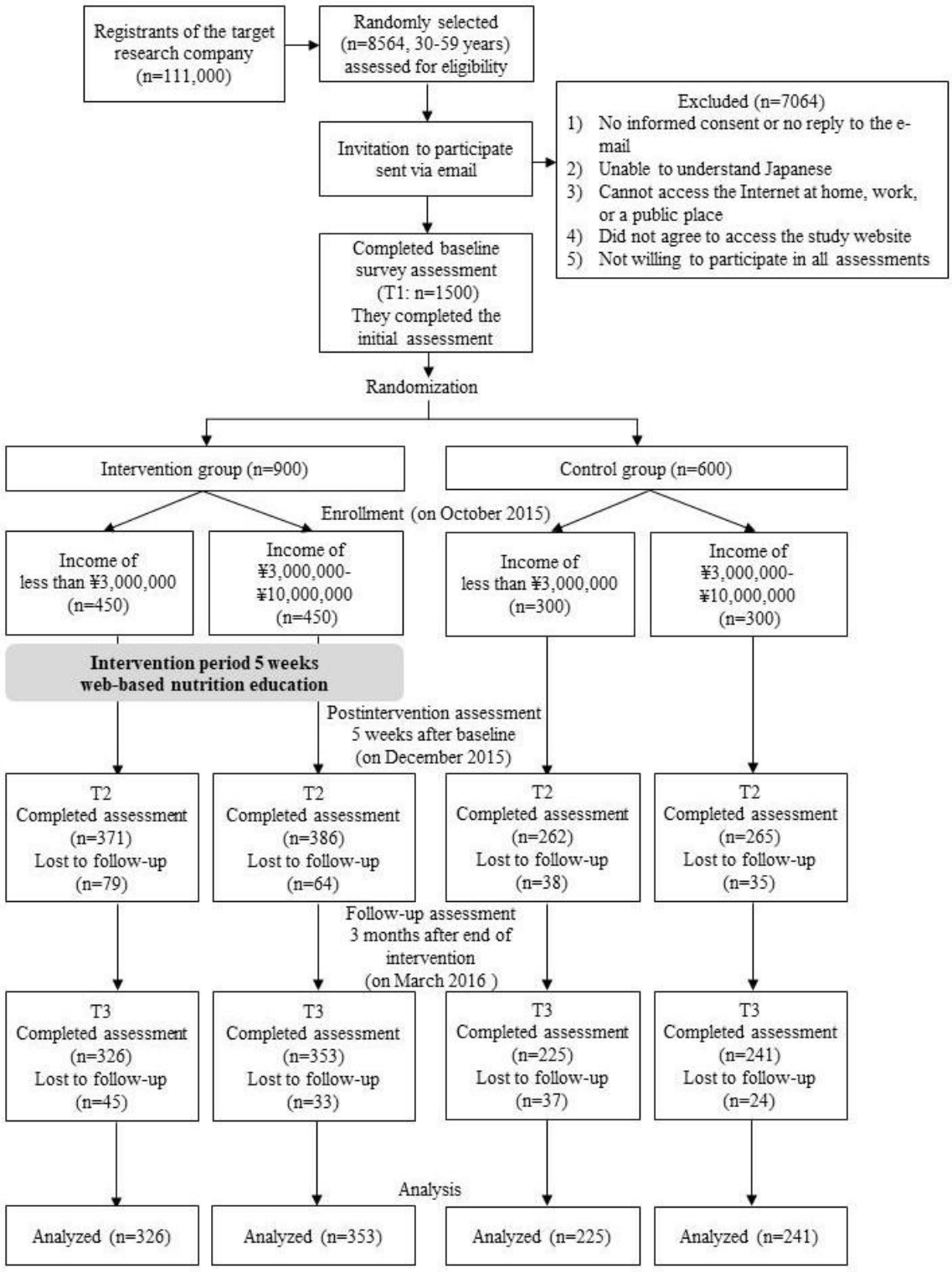
Trial Design and Ethics

We previously reported the details of the nutrition education program in a study protocol [30]. This study was a matched-design, randomized controlled trial (RCT). Participants were assessed by self-report at three time points: baseline, postintervention, and follow-up at 3 months. The study period ranged from October 2015 to March 2016. We obtained baseline data in October 2015 and postintervention data in December 2015; the follow-up period was March 2016 (ie, 3 months after postintervention). All intervention group participants completed the intervention in the same 5-week period. All control group participants completed the survey at all three time points, but did not undergo the intervention program. The RCT was approved by the Ethics Review Committee on Research with Human Subjects of Waseda University, Japan (2015-167), and Current Controlled Trials (UMIN-ICDR UMIN000019376).

Participants and Recruitment

Figure 1 shows the study participant recruitment and flow. A Japanese online research service company containing data from approximately 111,000 people (as of September 2015) conducted the survey at all three time points (baseline to follow-up). The research service company randomly selected 8564 adults aged 30 to 59 years to match the gender and age [31] distributions of Japan at baseline. We targeted adults aged 30 to 59 years because we felt that both the promotion of healthy eating and reduction in health disparities were particularly important in this group. In the past, we carried out a cross-sectional study on the relationship between socioeconomic status and dietary habits in this age group [14,30,32,33]. If participants met any of the exclusion criteria, they were not sent an email. Therefore, it is unknown why participants were excluded. The exclusion criterion were an annual income of more than ¥10,000,000 (this was equivalent to approximately US $83,333, in October 2015, US $1 was equivalent to approximately ¥120; 88.4% of the total population has an income of less than ¥10,000,000).

Figure 1. Flowchart showing participant recruitment, randomization, and evaluation of the Diet and Exercise Practices Project study.



Recruitment was terminated when the number of participants who agreed to participate reached 1500. The research service company randomly divided participants into intervention and control groups, and collected data via computer. The authors were blinded to the randomization. Participants received a detailed explanation of the research because of ethical considerations and were informed that they had been randomly assigned to their group. However, because participants did not obtain any information about the other participants, we believe that there was no contamination bias. The details of the incentives of this research are described in the study protocol [30].

The sample size was calculated using an effect size of .5, an alpha of .05, and power of .95 [30]. Among participants with incomes of less than ¥3,000,000 and those with incomes of ¥3,000,000 to ¥10,000,000, allocation was as follows: n=450 (intervention) and n=300 (control). Most adults in Japan have

incomes of ¥2,000,000 to ¥3,000,000, accounting for one-third of the Japanese population [34]. Our previous survey showed that the percentage of people eating 350 g (5 servings) of vegetables daily among individuals earning less than ¥3,000,000 was less (men: 5.5%; women: 10.4%) compared to those earning more than ¥3,000,000 [14]. Therefore, ¥3,000,000 was used as the relative cutoff point. Because most of the total population earn less than ¥10,000,000, this upper limit was set in consideration of ceiling effects [34]. The size of the control group was set at 600 participants; the expected dropout rate was approximately 50% according to the research service company during the survey period. The size of the intervention group was set at 900 participants, with an expected dropout rate of two-thirds. We also referred to the dropout percentage (15.3%) in Kothe et al [26] (the intervention period—30 days—was about the same as ours).

Procedure

The intervention group received emails (approximately 200 words in Japanese) with health information once a week on Monday between 6:00 am and 7:00 am. The email contained the following information: "website update announcement," "previous overview," "this summary," and "how to proceed with the site." For example, for step 2 of the intervention (which took place on the second week of the intervention), the email contained the following information:

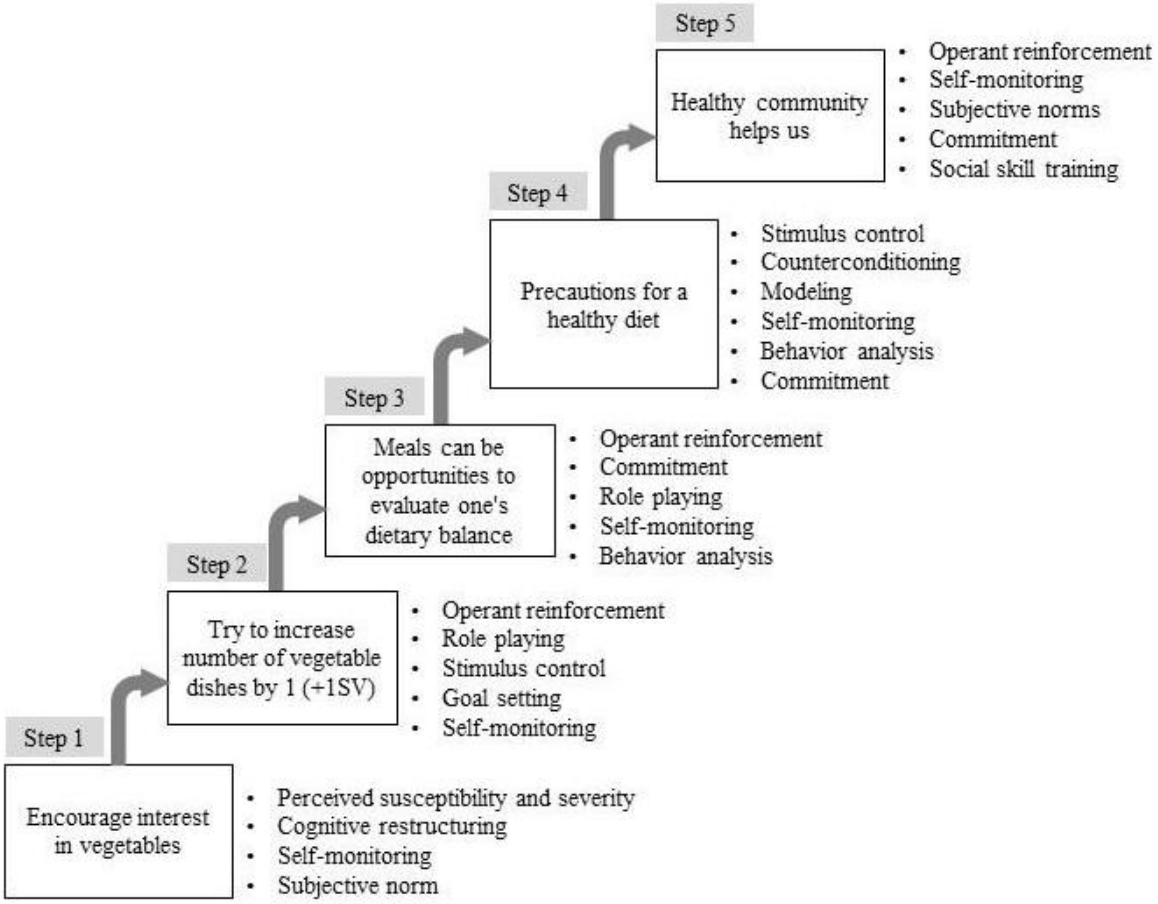
> *Hello, let us look back on your own eating habits is the first step toward health promotion. Step 2 has been updated so we will contact you. i) Diet: Review of Step 1 "How many vegetables dishes (servings) did you eat per day? Let's self-check and see" ii) Diet: Contents of Step 2 "Let's choose one more vegetable dish," iii) Please see 4 pages of each step in this order, 1) Today's points → 2) Do you know? → 3) Easy to devise → 4) Let's try it! Please look for evident information and let's choose what you can do.*

After completion of the 5-week intervention, participants received an email about the postintervention survey. Finally, participants received an email about the 3-month follow-up survey.

Intervention Program

An interactive website called the "Diet and Exercise Practices Project" [35] was developed. This is a free website that provides information, three monitoring sheets, and advice about healthy diets, increasing vegetable intake, and preventing lifestyle-related diseases. We hypothesized that achieving an approximately 70 g (1 serving) increase in vegetable intake might help lower-income adults "catch up" in intake compared to middle-income groups, while simultaneously contributing to the partial resolution of the overall deficient vegetable intake in Japanese adults.

Figure 2. Five steps and behavioral modification techniques of the nutrition education intervention program. The Web intervention period was 5 weeks.



The program consisted of a total of 20 pages of content, divided into five steps (one step contained four pages). The webpage was updated with one step every week. The program was based on the transtheoretical model (TTM) (Figure 2). Details of the program's theoretical framework are reported in a previously published study protocol [30]. In step 1, we used the health belief model to encourage movement from the precontemplative to the contemplative phase. In steps 2 and 3, social cognitive theory and the theory of planned behavior were used to encourage movement from the contemplative to the preparation phase. In step 4, social cognitive theory and the theory of planned behavior were again used, but this time to encourage movement from the preparation to the action phase. Finally, in step 5, strengthening of social networks and social support were used to promote a shift to the maintenance phase.

The four pages in each step were structured as follows: (1) "Today's point" (including a review of the previous week from the second week onward), which served as practical content; (2) "Do you know?" and (3) "Easy to devise," which were summaries; and (4) "Let's try it!" which involved supporting behavior change by using a worksheet. Figure 3 shows an example of the content on one page (ie, page 2 for step 2).

The control group surveys took place over the same period as the intervention group surveys. Control group participants received an email from the survey company informing them that they had been randomly assigned to a control group after the baseline. After a 5-week interval, participants received an email requesting them to take part in the postintervention survey. Three months later, the participants received an email requesting them to participate in a follow-up survey.

Figure 3. Example of content of the Web-based nutrition intervention program (1 page).

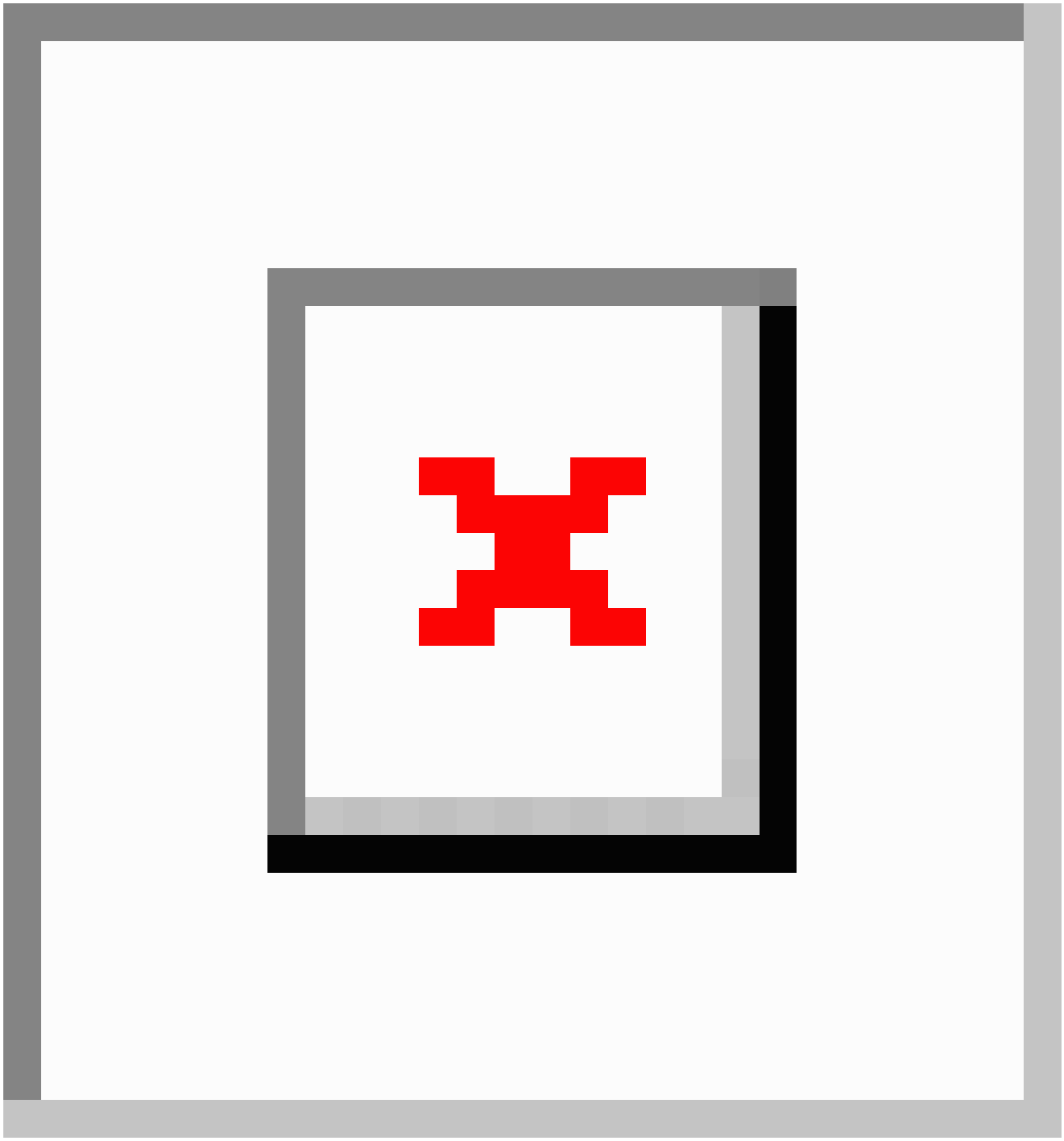


Table 1. Questionnaires and answers on vegetable intake at three time points.

Variable and questionnaire	Answer category
Vegetable intake	
How many vegetable dishes do you usually eat (dishes with vegetables as the main ingredient) per day? One dish is about one small bowl (70 g)	Number of dishes (servings/day)
Vegetable eating behavior (per week)[a]	
The following questions are about your normal dietary habits. Do you eat ample amounts of vegetables (5 small bowls/day, about 350 g)?	1=Almost every day; 2=4-5 days/week; 3=2-3 days/week; 4=almost none
Transtheoretical model[a]	
Which of the following matches your current dietary condition? Do you eat ample amounts of vegetables (5 small bowls/day, about 350 g)?	1=Maintenance (I have continued to eat them for more than 6 months); 2=action (I have continued to eat them for less than 6 months); 3=preparation (I sometimes eat or intend to eat within the next 30 days); 4=contemplation (although I do not currently eat them, I intend to start eating them within the next 6 months); 5=precontemplation (I do not eat them and I do not intend to start eating them within the next 6 months)
Perceived behavioral control[a]	
Do you believe you can do the following things to maintain your health, and your future health, with confidence? Do you have confidence in eating adequate amounts of vegetables (5 small dishes/day, or about 350 g)?	1=A lot of confidence (I have a lot of confidence in eating); 2=quite a lot of confidence (I have quite a lot of confidence in eating); 3=a little confidence (I have a little confidence in eating); 4=not a lot of confidence (I do not have a lot of confidence in eating); 5=very little confidence (I have very little confidence in eating); 6=not have any confidence (I do not have any confidence in eating)
Knowledge[a]	
Did you know that the recommended vegetable intake for maintaining health in adults is 350 g per day?	1=Yes; 2=no

[a]Participants chose one answer that best applied to them.

Data Collection and Outcome Measures

Participants were assessed via self-report at three time points: baseline, postintervention, and follow-up at 3 months. We have listed the details of the assessment items in Table 1. This study evaluated vegetable intake as the main outcome to assess the effectiveness of the nutrition education program. Ozawa et al [36] suggested that the number of vegetable dishes consumed may be a simpler and more valid measure of vegetable intake compared to a dietary record for both men and women. We presented participants with photographic examples of vegetable dishes (including the size and weight) before they answered the questionnaire. We referred to "The Japanese Food Guide Spinning Top" [37], wherein a dish where vegetables were the main ingredients (70 g) represented one serving.

Moreover, we performed a process evaluation of behavior change using various other outcomes, including vegetable eating behavior (per week) [38], stages of change, perceived behavioral control, and knowledge [39]. We used perceived behavioral control because it is an important concept [40] in behavior change. For the knowledge item, we showed photographic examples of vegetable dishes (including size and weight) before participants answered. Demographic variables included sex, age, marital status, number of people at home, employment status, and educational status.

Statistical Analyses

We compared the groups in terms of baseline sociodemographic characteristics using chi-square and Mann-Whitney *U* tests. Moreover, differences in baseline vegetable intake between participants and dropouts were assessed using unpaired *t* tests and one-way analyses of variance (ANOVAs). Amount of change in vegetable intake was analyzed using a general linear model. The mean change in vegetable intake was analyzed using Bonferroni-corrected comparisons following one-way ANOVAs for the different combinations of groups and time points. We compared the intervention effect on vegetable intake by group and time using two-way ANOVAs, and calculated the effect sizes (η^2). Other outcomes concerning vegetables were tested using McNemar test and the Wilcoxon signed rank test. The effects of multiple comparison were adjusted for using Bonferroni corrections. Participants lost to follow-up, that is those who did not complete the postintervention (n=216) or follow-up (n=139) surveys or who were otherwise missing any outcome data, were excluded from the analyses. This resulted in the exclusion of 355 of 1500 participants (23.67% of those randomly assigned) at baseline. Such an approach is in line with the revised CONSORT guidelines [41], as there are criticisms of and potential bias caused by imputing missing outcome data required for an intention-to-treat analysis. It has been pointed out that when the dropout rate is high, researchers should be cautious about conducting an intention-to-treat analysis. Indeed, in another RCT [42], an intention-to-treat analysis was not carried out because of a high dropout rate (14.9%). Therefore,

our analyses were not strictly intention to treat. All statistical analyses were performed using IBM SPSS Statistics version 21.0. A *P* value of .05 was used as the level of significance.

Results

Baseline Data

Table 2 shows baseline data collected from 1145 participants. The number of participants who completed the intervention in the intervention group (n=679) was as follows: low income (n=326, 72.4%) and middle income (n=353, 78.4%). In the control group (n=466), the number of participants who completed all three surveys was as follows: low income (n=225, 75.0%) and middle income (n=241, 80.3%). There were no differences in characteristics between the intervention group and the control group in either income group, except for marital status and number of people at home in the <¥3,000,000 group. Comparison of baseline characteristics between participants who were excluded and those who were included yielded the following differences: gender (included: men 596/1145, 52.1%, women 549/1145, 47.9%; excluded: men 154/355, 43.4%, women 201/355, 56.6%; *P*=.005) and educational status (included: junior high/high school 304/1134, 26.8%, 2-year college 297/1134, 26.2%, 4-year college/graduate school 533/1134, 47.0%; excluded: junior high/high school 114/348 32.8%, 2-year college 99/1134 28.4%, 4-year college/graduate school 135/1134, 38.8%; *P*=.005).

Table 2. Baseline sociodemographic characteristics of the study participants by income level (N=1145).

Variable	<¥3,000,000[a]			¥3,000,000-¥10,000,000[a]		
	Intervention (n=326) n (%)	Control (n=225) n (%)	*P*	Intervention (n=353) n (%)	Control (n=241) n (%)	*P*
Gender[b]			.80			.93
Men	169 (51.8)	120 (53.3)		183 (51.8)	124 (51.5)	
Women	157 (48.2)	105 (46.7)		170 (48.2)	117 (48.5)	
Age (years)[c]			.39			.23
30-39	106 (32.5)	76 (33.8)		117 (33.1)	73 (30.3)	
40-49	116 (35.6)	88 (39.1)		138 (39.1)	89 (36.9)	
50-59	104 (31.9)	61 (27.1)		98 (27.8)	79 (32.8)	
Marital status[b]			.04			.47
Not married	217 (66.6)	169 (75.1)		116 (32.9)	72 (29.9)	
Married	109 (33.4)	56 (24.9)		237 (67.1)	169 (70.1)	
Number of people at home[b,d]			.03			.35
≥2	198 (66.4)	111 (56.6)		292 (85.6)	194 (82.6)	
1	100 (33.6)	85 (43.4)		49 (14.4)	41 (17.4)	
Employment status[b,d]			.39			.60
Not employed	103 (32.5)	61 (28.6)		74 (21.4)	46 (19.2)	
Employed	214 (67.5)	152 (71.4)		272 (78.6)	193 (80.8)	
Educational status[c,d]			.06			.91
Junior high/high school	114 (35.5)	72 (32.3)		71 (20.3)	47 (19.6)	
2-year college	92 (28.7)	47 (21.1)		93 (26.6)	65 (27.1)	
4-year college/graduate school	115 (35.8)	104 (46.6)		186 (53.1)	128 (53.3)	

[a]¥120=US $1 (October 2015).

[b]Chi-square test.

[c]Mann-Whitney *U* test.

[d]Percentage excludes unknown/other answers. In <¥3,000,000: number of people at home (n=28), employment status (n=9), educational status (n=5) in the intervention group; number of people at home (n=29), employment status (n=12), educational status (n=2) in the control group. In ¥3,000,000-¥10,000,000: number of people at home (n=12), employment status (n=7), educational status (n=3) in the intervention group; number of people at home (n=6), employment status (n=2), educational status (n=1) in the control group.

Outcomes

Means and Mean Differences in Vegetable Intake at Each Time Point

Table 3 shows the mean (SD) vegetable intake at each time point. The participants with low income at baseline in the intervention group showed a lower vegetable intake compared with middle-income participants in both the intervention and control groups. The same pattern was found for participants with low income at baseline in the control group. There were no other differences between the groups. We confirmed that there were no differences in baseline vegetable intake between participants who were included in the analysis and those who dropped out (*P*=.91). The mean difference in vegetable intake at postintervention also increased in low-income intervention group participants compared to baseline (mean 0.42 servings, 95% CI 0.11 to 0.72, *P*<.001). In the control group among low-income participants, the mean vegetable intake at postintervention was not much different from that at baseline (mean 0.05 servings, 95% CI –0.26 to 0.36); the difference between follow-up and baseline was also minor (mean 0.03 servings, 95% CI –0.28 to 0.34). For middle-income participants, the mean vegetable intake at postintervention was barely different from that baseline (mean 0.04 servings, 95% CI –0.27 to 0.36); the same was true comparing follow-up and baseline (mean 0.03 servings, 95% CI –0.29 to 0.34).

Effect Size of Vegetable Intake

Table 4 shows the effect size of vegetable intake by income. Two-way ANOVAs showed that both low- and middle-income participants had significant main effects of group and time, and a significant interaction. Multiple comparisons (Figure 4) showed that vegetable intake among low-income participants increased between baseline and postintervention. Although it did not decrease significantly between postintervention and follow-up, the difference between baseline and follow-up was not significant. There were no changes in vegetable intake among middle-income participants when comparing any time point (Figure 5). However, multiple comparisons revealed that vegetable intake among low-income participants at baseline was lower than that among middle-income participants (baseline: *P*=.003). At postintervention and follow-up, the difference between income groups had disappeared (postintervention: *P*=.16; follow-up: *P*=.045).

Table 3. Mean (SD) and mean difference (95% CI) in each vegetable intake measure at the three time points.

Time	<¥3,000,000[a] (servings/day)				¥3,000,000-¥10,000,000[a] (servings/day)			
	Intervention (n=326) mean (SD)[b]	Control (n=225) mean (SD)[b]	Mean difference (95% CI)[c]	*P*	Intervention (n=353) mean (SD)[b]	Control (n=241) mean (SD)[b]	Mean difference (95% CI)[c]	*P*
Baseline	2.08 (1.49)	1.88 (1.38)			2.42 (1.50)	2.44 (1.40)		
Postintervention	2.50 (1.79)	1.93 (1.37)	0.42 (0.11, 0.72)	<.003	2.67 (1.46)	2.49 (1.49)	0.25 (–0.02, 0.52)	.08
Follow-up	2.23 (1.54)	1.91 (1.33)	0.15 (–0.15, 0.46)	.67	2.47 (1.46)	2.47 (1.44)	0.04 (–0.23, 0.31)	>.99

[a]¥120=US $1 (October 2015).

[b]Mean (SD) of servings/day. General linear model. One-way ANOVA *P*<.05, Bonferroni-corrected post hoc comparisons using *t* test. Significance was based on *P*<.05/6=.008.

[c]Mean difference (95% CI) in servings/day in intervention group from baseline.

Table 4. The effect size (η^2)[a] in comparisons of vegetable intake between groups and times using two-way repeated-measures ANOVAs[b].

Items	<¥3,000,000[c]			¥3,000,000-¥10,000,000[c]		
	Partial η^2	η^2	*P*	Partial η^2	η^2	*P*
Group	0.10	0.04	.03	0.00	0.00	.62
Time	0.01	0.01	.02	0.01	0.01	.006
Group*time	0.10	0.01	.009	0.01	0.01	.046

[a]Effect size (low: η^2=0.01; middle: η^2=0.06; high: η^2=0.14).

[b]General linear model adjusted for baseline marital status and number of people at home in <¥3,000,000. Dependent variable: vegetable intake servings.

[c]In <¥3,000,000 group, self-reported vegetable intake at all three time points: n=326 (intervention) and n=225 (control). In ¥3,000,000-¥10,000,000 group, self-reported vegetable intake at all three time points: n=353 (intervention) and n=241 (control). ¥120=US $1 (October 2015).

Figure 4. In the <¥3,000,000 group, means of self-reported vegetable intake at baseline, postintervention, and follow-up at 3 months in the intervention group (solid line) and control group (dotted line). Data were analyzed by using two-way ANOVAs. *General linear model, significance was based on P <.05/3=.02 (Bonferroni-corrected).

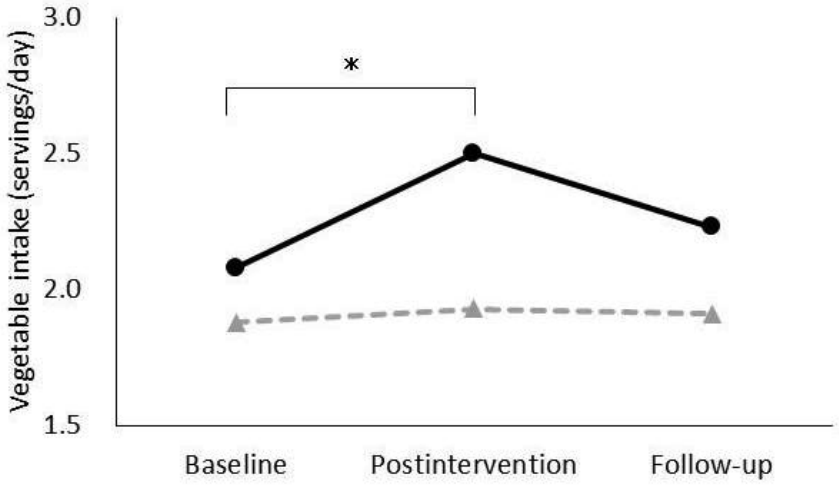
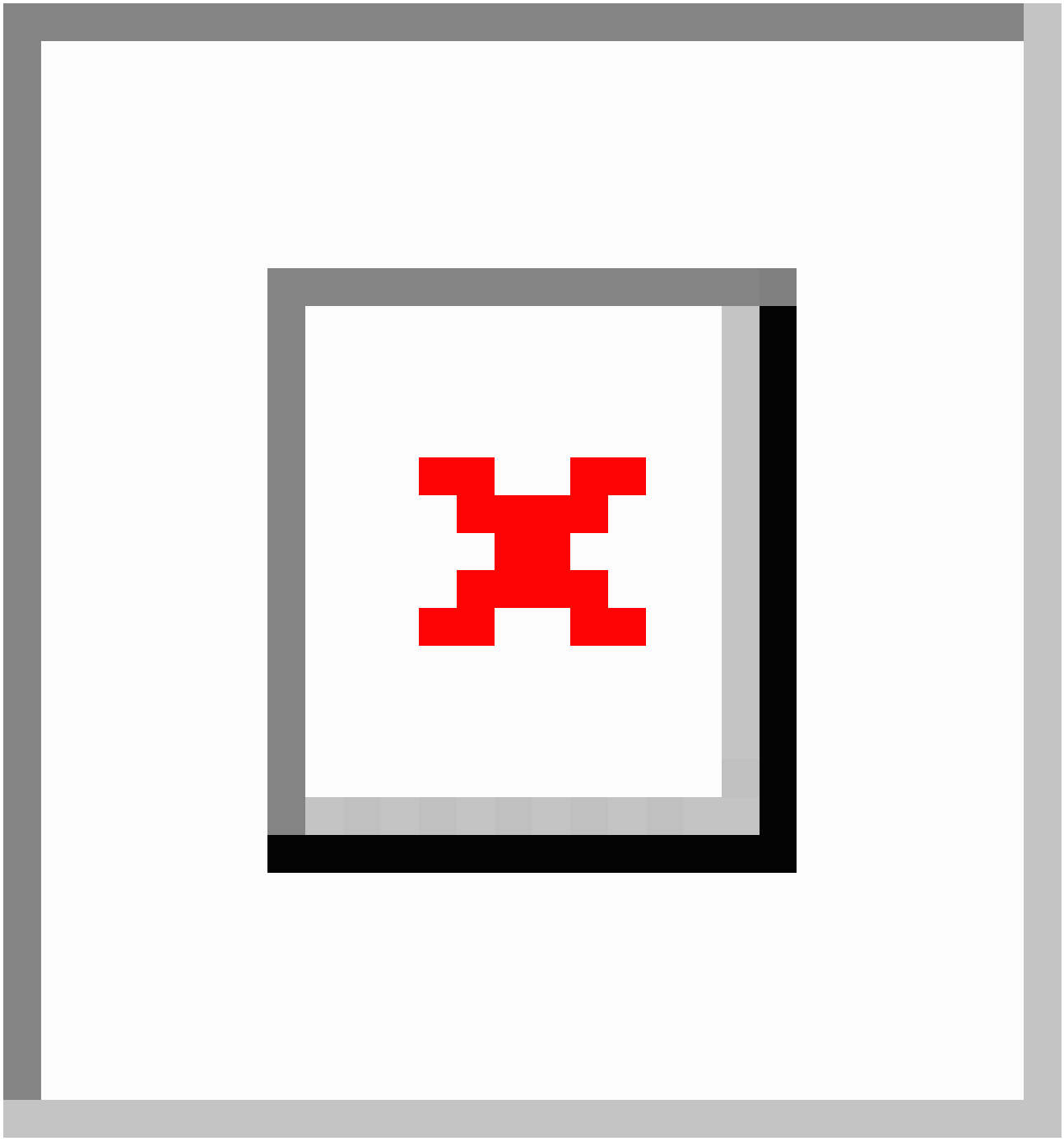


Figure 5. In the ¥3,000,000-¥10,000,000 group, means of self-reported vegetable intake at baseline, postintervention, and follow-up at 3 months in the intervention group (solid line) and the control group (dotted line). Data were analyzed by using two-way ANOVAs.



Results for Other Vegetable Intake Variables

Tables 5 and 6 show the results for eating vegetables behavior, stage of change, perceived behavioral control, and knowledge of vegetable intake. Low-income participants (<¥3,000,000) in the intervention group showed improvements in eating vegetables, stages of change, perceived behavioral control, and knowledge at postintervention compared to baseline. Furthermore, the improvements in eating vegetables and dietary knowledge were maintained between postintervention and follow-up. In middle-income participants (¥3,000,000-¥10,000,000), only improvements in knowledge were maintained from baseline to postintervention, and from baseline to follow-up. However, in the control group, improvements in knowledge were maintained from baseline to postintervention and from baseline to follow-up among both income groups (all P<.001).

Table 5. Baseline (T1), postintervention (T2), and follow-up (T3) in change in behavior, transtheoretical model, perceived behavioral control, and knowledge of vegetable intake as a result of Web-based intervention among adults with an income of <¥3,000,000.[a]

Variable	Intervention (n=326)					Control (n=225)				
	T1 n (%)	T2 n (%)	T3 n (%)	T1-T2 P^b	T1-T3 P^b	T1 n (%)	T2 n (%)	T3 n (%)	T1-T2 P^b	T1-T3 P^b
Eating vegetable behavior[c]				<.001	.008				.13	.005
Almost every day	24 (7.4)	39 (12.0)	32 (9.8)			14 (6.2)	14 (6.2)	17 (7.6)		
4-5 days/week	36 (11.0)	52 (16.0)	38 (11.7)			18 (8.0)	21 (9.3)	30 (13.3)		
2-3 days/week	91 (27.9)	97 (29.8)	109 (33.4)			75 (33.3)	79 (35.1)	61 (27.1)		
Almost none	175 (53.7)	138 (42.3)	147 (45.1)			118 (52.4)	111 (49.3)	117 (52.0)		
Transtheoretical model[d]				<.001	.27				.10	.29
Maintenance	56 (17.2)	74 (22.7)	65 (19.9)			27 (12.0)	34 (15.1)	40 (17.8)		
Action	17 (5.2)	16 (4.9)	19 (5.8)			10 (4.4)	19 (8.4)	9 (4.0)		
Preparation	108 (33.1)	123 (37.7)	100 (30.7)			82 (36.4)	62 (27.6)	64 (28.4)		
Contemplation	84 (25.8)	75 (23.0)	83 (25.5)			62 (27.6)	68 (30.2)	69 (30.7)		
Precontemplation	61 (18.7)	38 (11.7)	59 (18.1)			44 (19.6)	42 (18.7)	43 (19.1)		
Perceived behavioral control[c]				<.001	.06				.32	.25
A lot of confidence	15 (4.6)	24 (7.4)	16 (4.9)			10 (4.4)	9 (4.0)	8 (3.6)		
Quite a lot of confidence	30 (9.2)	33 (10.1)	38 (11.7)			10 (4.4)	18 (8.0)	23 (10.2)		
A little confidence	49 (15.0)	56 (17.2)	54 (16.6)			33 (14.7)	32 (14.2)	33 (14.7)		
Not a lot of confidence	108 (33.1)	118 (36.2)	109 (33.4)			78 (34.7)	72 (32.0)	75 (33.3)		
Very little confidence	51 (15.6)	42 (12.9)	41 (12.6)			40 (17.8)	40 (17.8)	29 (12.9)		
Not any confidence	73 (22.4)	53 (16.3)	68 (20.9)			54 (24.0)	54 (24.0)	57 (25.3)		
Knowledge				<.001	<.001				<.001	<.001
Yes	106 (32.5)	174 (53.4)	177 (54.3)			60 (26.7)	90 (40.0)	93 (41.3)		
No	220 (67.5)	152 (46.6)	149 (45.7)			165 (73.3)	135 (60.0)	132 (58.7)		

[a]¥120=US $1 (October 2015).

[b]Ordinal scale: Wilcoxon signed rank test, Bonferroni-corrected post hoc comparisons made using Mann-Whitney *U* test (P<.05/3=.02). Nominal scale: McNemar test, Bonferroni-corrected post hoc comparisons made using chi-square test (P<.05/3=.02).

[c]The percentage might not reach 100% in some cases because the rate was rounded off.

[d]Transtheoretical model (TTM) 5 stages of change: (1) maintenance (I have continued to eat them for more than 6 months); (2) action (I have continued to eat them for less than 6 months); (3) preparation (I sometimes eat them or intend to eat them within the next 30 days); (4) contemplation (although I do not eat them currently, I intend to start eating them within the next 6 months); (5) precontemplation (I do not eat them and I do not intend to start eating them within the next 6 months).

Table 6. Baseline (T1), postintervention (T2), and follow-up (T3) in change in behavior, transtheoretical model, perceived behavioral control, and knowledge about vegetable intake in this Web-based intervention among adults with an income of ¥3,000,000-¥10,000,000.[a]

Variable	Intervention (n=353)					Control (n=241)				
	T1 n (%)	T2 n (%)	T3 n (%)	T1-T2 P^b	T1-T3 P^b	T1 n (%)	T2 n (%)	T3 n (%)	T1-T2 P^b	T1-T3 P^b
Eating vegetable behavior[c]				.38	.12				.03	.02
Almost every day	32 (9.1)	33 (9.3)	34 (9.6)			19 (7.9)	19 (7.9)	18 (7.5)		
4-5 days/week	50 (14.2)	56 (15.9)	73 (20.7)			38 (15.8)	44 (18.3)	48 (19.9)		
2-3 days/week	125 (35.4)	134 (38.0)	120 (34.0)			86 (35.7)	103 (42.7)	100 (41.5)		
Almost none	146 (41.4)	130 (36.8)	126 (35.7)			98 (40.7)	75 (31.1)	75 (31.1)		
Transtheoretical model[c,d]				.25	.14				.07	.24
Maintenance	87 (24.6)	81 (22.9)	77 (21.8)			49 (20.3)	64 (26.6)	52 (21.6)		
Action	23 (6.5)	24 (6.8)	42 (11.9)			12 (5.0)	15 (6.2)	21 (8.7)		
Preparation	106 (30.0)	149 (42.2)	128 (36.3)			96 (39.8)	84 (34.9)	91 (37.8)		
Contemplation	94 (26.6)	72 (20.4)	64 (18.1)			61 (25.3)	48 (19.9)	54 (22.4)		
Precontemplation	43 (12.2)	27 (7.6)	42 (11.9)			23 (9.5)	30 (12.4)	23 (9.5)		
Perceived behavioral control[c]				.58	.15				.04	.95
A lot of confidence	19 (5.4)	26 (7.4)	19 (5.4)			13 (5.4)	18 (7.5)	8 (3.3)		
Quite a lot of confidence	53 (15.0)	44 (12.5)	47 (13.3)			33 (13.7)	36 (14.9)	31 (12.9)		
A little confidence	71 (20.1)	71 (20.1)	64 (18.1)			45 (18.7)	39 (16.2)	47 (19.5)		
Not a lot of confidence	122 (34.6)	132 (37.4)	135 (38.2)			81 (33.6)	94 (39.0)	100 (41.5)		
Very little confidence	48 (13.6)	47 (13.3)	43 (12.2)			45 (18.7)	36 (14.9)	33 (13.7)		
Not any confidence	40 (11.3)	33 (9.3)	45 (12.7)			24 (10.0)	18 (7.5)	22 (9.1)		
Knowledge				<.001	<.001				<.001	<.001
Yes	109 (30.9)	186 (52.7)	191 (54.1)			69 (28.6)	90 (37.3)	110 (45.6)		
No	244 (69.1)	167 (47.3)	162 (45.9)			172 (71.4)	151 (62.7)	131 (54.4)		

[a]¥120=US $1 (October 2015).

[b]Ordinal scale: Wilcoxon signed rank test, Bonferroni-corrected post hoc comparisons made using Mann-Whitney *U* test ($P<.05/3=.02$). Nominal scale: McNemar test, Bonferroni-corrected post hoc comparisons made using chi-square test ($P<.05/3=.02$).

[c]The percentage might not reach 100% in some cases because the rate was rounded off.

[d]Transtheoretical model (TTM) 5 stages of change: (1) maintenance (I have continued to eat them for more than 6 months); (2) action (I have continued to eat them for less than 6 months); (3) preparation (I sometimes eat them or intend to eat them within the next 30 days); (4) contemplation (although I do not eat them currently, I intend to start eating them within the next 6 months); (5) precontemplation (I do not eat them and I do not intend to start eating them within the next 6 months).

Discussion

Principal Results

A strength of this study was its RCT design and stratification by income to investigate the reduction in vegetable intake disparity. The main finding was that vegetable intake and related processes among low-income participants improved, thus reducing the existing disparities with the middle-income group. We suggest that this Web-based nutrition education program based on multiple health behavior theories is an effective intervention for low-income adults. To our knowledge, this is the first Web-based intervention study to investigate reductions in vegetable intake disparities in adults. Further improvements in the intervention program are necessary to increase intake and maintain that increase throughout a follow-up period among middle-income adults.

The vegetable intake among low-income participants increased by mean 0.42 servings (95% CI 0.11-0.72) after the intervention, which helped reduce the vegetable intake disparity between incomes. Additionally, behavioral change processes such as dietary behavior, stages of change, self-efficacy, and knowledge improved. Most past Web-based studies were conducted outside Japan. Bensley et al [24] reported an increase of 0.59 servings after a Web-based nutrition education for 6 months with adults, Sternfeld et al [23] an increase of 0.18 servings using an email intervention for 16 weeks with employees, and Kothe et al [26] an increase of 0.84 servings using an email intervention for 30 days with undergraduate students. Thus, in all three of these studies, the intervention led to increased fruit and vegetable

intake. However, these studies evaluated vegetables and fruit within the same category. The Japanese Food Guide Spinning Top, a Japanese food guide, classifies vegetables and fruits in different categories [37]. This study showed a clearly positive intervention effect for vegetable intake only. Past studies in Japan showed that nutrition education interventions increased vegetable intake by 0.32 servings in 24 weeks among male workers [43], and by 0.30 servings 1 year later in employees [44]. The improvements in vegetable intake and improved behavioral change processes are further strengths of this study. Our results contributed to reducing disparities in vegetable intake between low- and middle-income individuals by using a very short-term (5-week) Web-based intervention.

The nutrition education program also has two important strengths. First, the program was based on the stages of change, which are thought to be applicable to nutrition interventions [45]. Many previous studies have found support for methods based on multiple health behavior theories aimed at increasing vegetable intake [46,47]. In this study, participants were assumed to be in the precontemplation stage baseline (Figure 2). Because improved self-efficacy is essential for behavior change, we made sure that all steps of the program focused on improving self-efficacy. We expect that the composition of this program helped increase vegetable intake among low-income individuals, who may have had low self-efficacy (perceived behavioral control).

Second, Park et al [22] reported that 88% of participants completed their 30-day Web intervention for adults. Kothe et al reported 85% [26] and 80% [27] of participants completed the intervention with a 30-day email intervention for undergraduate students. This study had a roughly equivalent number of participants who completed the intervention (low income: 82.4%; middle income: 85.7% at postintervention). This is possibly because participants received an email including a weekly summary of the program and informative support. Additionally, the intervention was highly accessible (eg, time, place, situation) because participants could complete the activities using their mobile phone or personal computer. There are extremely few previous reports on Web-based interventions for Japanese adults [48]; as such, our Web-based nutrition education program is not only highly effective, but also provides novel evidence.

The improvement in vegetable intake for low-income participants in the intervention group disappeared by the 3-month follow-up. This might have been because we used only one theoretical approach (see step 5) to promote the change from the action to maintenance stage. Therefore, it is necessary to strengthen this aspect of the intervention. For example, Japanese traditional food culture distinguishes between seasonal dishes and foods for all four seasons. We could distribute nutrition information and recipes about seasonal foods such as vegetables during a follow-up period to promote continued vegetable intake. In addition, we could regularly tweet reviews of the program content, and send reminders of the effort needed to prevent reversal of behavior change. Behavior change can be regarded as habitual if it is maintained for more than 6 months. After observing the program-related improvements in this study, it is worth attempting these approaches during follow-up to maintain the intervention effect.

Vegetable intake among middle-income participants might not have increased after the intervention because of the program's use of an inappropriate approach to behavior change for this income group. For example, food access and perceptions of the food environment might differ depending on income [49]. Perceived behavioral control, which was the focus of this program [30], has been found to be low among low-income women [20]. Thus, a program focusing on perceived behavioral control might have promoted ingestion of vegetables up to a certain amount. For further improvement, it is necessary to identify the factors affecting vegetable intake according to income and develop more appropriate intervention methods.

The reason for the low effect sizes was probably the smaller variety of content and shorter intervention periods than in previous studies. The effect size of a 30-day nutrition education intervention by email in undergraduate students [26] was roughly the same as in this study. However, Alexander et al [25] found medium effect sizes when using a website targeting adults. Their program had a rich variety of content, such as food education using a short video and audio files, and presented 300 fruit- and vegetable-based recipes. Furthermore, the intervention period was one year, which was considerably longer than was ours. The content of our program is a website of approximately 20 pages containing information and worksheets combining text and images. We expect larger effects if we increased the variety of the content and the length of the intervention period.

Limitations

Some limitations warrant discussion. First, we were careful to extract samples matching the Japanese demographic distribution. Nonetheless, our participants had a higher education level compared to the census [50]. Second, we could not identify the factors that improved the control group's knowledge. Possibly, they acquired the knowledge during the survey or they were exposed to health promotion strategies elsewhere. However, the results show that behavioral change does not occur merely through improving knowledge. Third, the design was not strictly intention to treat. Using an approach to impute missing outcome data for the relatively large number of dropouts (23.6% of the sample) can cause potential biases. We analyzed them in comparison to their originally assigned group, and confirmed that there were no baseline differences in vegetable intake between the participants included in the analysis and the dropouts. Fourth, regrettably, we have no data on the weight status, health status, or chronic diseases of participants. We did not assess body weight status because the validity of self-assessment of body weight is unknown. Furthermore, a diagnosis by a doctor is necessary to determine the presence or absence of a chronic disease. Exercise is currently being investigated by other project teams; therefore, we could not handle the data on exercise. Other relevant data, such as frequency of intake of other foods, should be examined in the future. This would help in generalizing the results of our study. Fifth, we could not investigate the relationship between the intervention dose and its effect. Regrettably, the website set a common password for all participants because we had

insufficient research funding and were unable to handle personal information such as individual ID and password settings. By developing apps or other tools in cooperation with companies in the future, we would be able to further develop this line of research. Finally, the results can only apply at this present time to Japanese individuals aged 30 to 59 years and with incomes less than ¥10,000,000, thus limiting the generalizability of the findings.

Practical Implications

This program has the following implications. The intervention succeeded in increasing vegetable intake without being restricted to a single geographical area. This shows the possibility that our nutrition education program can spread widely in the future. Furthermore, the program has a systematic composition, containing five steps of four pages of content each: (1) “Today’s point,” 2) “Do you know?” 3) “Easy to devise,” and 4) “Let’s try it!” It is worth investigating whether the program can achieve the same effect using other methods (eg, higher frequency of emails [20 times], face-to-face delivery of content). Further research might aim to clarify which components of Web-based interventions or the program framework contribute to reducing vegetable intake disparities.

Conclusion

The findings from this RCT indicate that this Web-based nutrition education program can increase vegetable intake among low-income adults, thus contributing to a reduction in vegetable intake disparities.

Acknowledgments

The authors would like to thank all participating agencies and professionals, as well as the participants. This research was supported by a Grant-in-Aid for Challenging Exploratory Research (JSPS KAKENHI Grant Number JP25560357) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Number JP16J02348).

Authors' Contributions

SN designed the study, analyzed the data, and drafted the paper. TI and TA provided advice about the study design, supervised, and helped draft the manuscript. KH helped draft the manuscript. All authors contributed to and have approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT‐EHEALTH checklist (V.1.6.1).

[PDF File (Adobe PDF File), 620KB - jmir_v19i11e377_app1.pdf]

References

1. Lantz PM, House JS, Lepkowski JM, Williams DR, Mero RP, Chen J. Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults. JAMA 1998 Jun 3;279(21):1703-1708. [Medline: 9624022]
2. Oldroyd J, Burns C, Lucas P, Haikerwal A, Waters E. The effectiveness of nutrition interventions on dietary outcomes by relative social disadvantage: a systematic review. J Epidemiol Community Health 2008 Jul;62(7):573-579. [doi: 10.1136/jech.2007.066357] [Medline: 18559438]
3. McGill R, Anwar E, Orton L, Bromley H, Lloyd-Williams F, O'Flaherty M, et al. Are interventions to promote healthy eating equally effective for all? Systematic review of socioeconomic inequalities in impact. BMC Public Health 2015 Sep 15;15:894. [doi: 10.1186/s12889-015-2162-y] [Medline: 26373840]
4. Darmon N, Drewnowski A. Does social class predict diet quality? Am J Clin Nutr 2008 May;87(5):1107-1117 [FREE Full text] [Medline: 18469226]
5. World Cancer Research Fund, American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective. Washington, DC: American Institute for Cancer Research; 2007.
6. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. Changes in diet and lifestyle and long-term weight gain in women and men. N Engl J Med 2011 Jun 23;364(25):2392-2404 [FREE Full text] [doi: 10.1056/NEJMoa1014296] [Medline: 21696306]
7. Nagura J, Iso H, Watanabe Y, Maruyama K, Date C, Toyoshima H, JACC Study Group. Fruit, vegetable and bean intake and mortality from cardiovascular disease among Japanese men and women: the JACC Study. Br J Nutr 2009 Jul;102(2):285-292. [doi: 10.1017/S0007114508143586] [Medline: 19138438]
8. Gillman MW, Cupples LA, Gagnon D, Posner BM, Ellison RC, Castelli WP, et al. Protective effect of fruits and vegetables on development of stroke in men. JAMA 1995 Apr 12;273(14):1113-1117. [Medline: 7707599]

9. Bazzano LA, Serdula MK, Liu S. Dietary intake of fruits and vegetables and risk of cardiovascular disease. Curr Atheroscler Rep 2003 Nov;5(6):492-499. [Medline: 14525683]
10. Ministry of Health, Labour and Welfare. The National Health and Nutrition Survey [in Japanese]. 2014. URL: http://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/0000117311.pdf [accessed 2017-10-19] [WebCite Cache ID 6uKZ9Aq1X]
11. Kondo N. Socioeconomic disparities and health: impacts and pathways. J Epidemiol 2012;22(1):2-6 [FREE Full text] [Medline: 22156290]
12. Kagamimori S, Gaina A, Nasermoaddeli A. Socioeconomic status and health in the Japanese population. Soc Sci Med 2009 Jun;68(12):2152-2160. [doi: 10.1016/j.socscimed.2009.03.030] [Medline: 19375838]
13. Ministry of Health, Labour and Welfare. Health Japan 21 (the second term). 2012. URL: http://www.mhlw.go.jp/file/06-Seisakujouhou-10900000-Kenkoukyoku/0000047330.pdf [accessed 2017-10-19] [WebCite Cache ID 6uKZU97g9]
14. Nakamura S, Inayama T, Arao T. Association of vegetable consumption with eating behaviors in Japanese adults: a cross-sectional analysis [in Japanese]. Jpn J Health Educ Promot 2016;24(2):65-80. [doi: 10.11260/kenkokyoiku.24.65]
15. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
16. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
17. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Processes 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
18. Thomson CA, Ravia J. A systematic review of behavioral interventions to promote intake of fruit and vegetables. J Am Diet Assoc 2012 Oct;111(10):1523-1535. [doi: 10.1016/j.jada.2011.07.013] [Medline: 21963019]
19. Appleton KM, Hemingway A, Saulais L, Dinnella C, Monteleone E, Depezay L, et al. Increasing vegetable intakes: rationale and systematic review of published interventions. Eur J Nutr 2016 Apr;55(3):869-896 [FREE Full text] [doi: 10.1007/s00394-015-1130-8] [Medline: 26754302]
20. Henry H, Reimer K, Smith C, Reicks M. Associations of decisional balance, processes of change, and self-efficacy with stages of change for increased fruit and vegetable intake among low-income, African-American mothers. J Am Diet Assoc 2006 Jun;106(6):841-849. [doi: 10.1016/j.jada.2006.03.012] [Medline: 16720125]
21. Bandura A. Self-Efficacy: The Exercise of Control. Basingstoke: Macmillan; 1997.
22. Park A, Nitzke S, Kritsch K, Kattelmann K, White A, Boeckner L, et al. Internet-based interventions have potential to affect short-term mediators and indicators of dietary behavior of young adults. J Nutr Educ Behav 2008;40(5):288-297. [doi: 10.1016/j.jneb.2008.02.001] [Medline: 18725147]
23. Sternfeld B, Block C, Quesenberry CPJ, Block TJ, Husson G, Norris JC, et al. Improving diet and physical activity with ALIVE: a worksite randomized trial. Am J Prev Med 2009 Jun;36(6):475-483. [doi: 10.1016/j.amepre.2009.01.036] [Medline: 19460655]
24. Bensley RJ, Anderson JV, Brusk JJ, Mercer N, Rivas J. Impact of internet vs traditional Special Supplemental Nutrition Program for Women, Infants, and Children nutrition education on fruit and vegetable intake. J Am Diet Assoc 2011 May;111(5):749-755. [doi: 10.1016/j.jada.2011.02.010] [Medline: 21515124]
25. Alexander GL, McClure JB, Calvi JH, Divine GW, Stopponi MA, Rolnick SJ, MENU Choices Team. A randomized clinical trial evaluating online interventions to improve fruit and vegetable consumption. Am J Public Health 2010 Feb;100(2):319-326. [doi: 10.2105/AJPH.2008.154468] [Medline: 20019315]
26. Kothe EJ, Mullan BA, Butow P. Promoting fruit and vegetable consumption. Testing an intervention based on the theory of planned behaviour. Appetite 2012 Jun;58(3):997-1004. [doi: 10.1016/j.appet.2012.02.012] [Medline: 22349778]
27. Kothe EJ, Mullan BA. A randomised controlled trial of a theory of planned behaviour to increase fruit and vegetable consumption. Fresh Facts. Appetite 2014 Jul;78:68-75. [doi: 10.1016/j.appet.2014.03.006] [Medline: 24656949]
28. Buller DB, Woodall WG, Zimmerman DE, Slater MD, Heimendinger J, Waters E, et al. Randomized trial on the 5 a day, the Rio Grande Way Website, a web-based program to improve fruit and vegetable consumption in rural communities. J Health Commun 2008 Apr;13(3):230-249 [FREE Full text] [doi: 10.1080/10810730801985285] [Medline: 18569356]
29. Neuenschwander LM, Abbott A, Mobley AR. Comparison of a web-based vs in-person nutrition education program for low-income adults. J Acad Nutr Diet 2013 Jan;113(1):120-126. [doi: 10.1016/j.jand.2012.07.034] [Medline: 23092741]
30. Nakamura S, Inayama T, Arao T. A randomized-controlled trial focusing on socio-economic status for promoting vegetable intake among adults using a web-based nutrition intervention programmetudy protocol. BMC Public Health 2017 Jan 13;17(1):74. [doi: 10.1186/s12889-016-3907-y] [Medline: 28086921]
31. Statistics Bureau Ministry of Internal Affairs and Communications. 2013. The population census of Japan [in Japanese] URL: http://www.soumu.go.jp/menu_news/s-news/01gyosei02_02000055.html [accessed 2017-10-19] [WebCite Cache ID 6uKaiL5pn]
32. Nakamura S, Inayama T, Hata K, Matsushita M, Takahashi M, Harada K, et al. Association of household income and education with eating behaviors in Japanese adults: a cross-sectional study. BMC Public Health Jan 2016 Jan 22;16:61. [doi: 10.1186/s12889-016-2748-z] [Medline: 26800891]

33. Nakamura S, Inayama T, Hata K, Matsushita M, Takahashi M, Harada K, et al. Association between health literacy and eating vegetables and socioeconomic status in adults [in Japanese]. Japan Soc Health Promot 2016;18(2):27-35.
34. Ministry of Health, Labour and Welfare. Comprehensive survey of living conditions [in Japanese]. 2012. URL: http://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa12/dl/03.pdf [accessed 2017-10-19] [WebCite Cache ID 6uKb2w3FO]
35. Diet and Exercise Practices Project [in Japanese]. URL: http://healthpromotionqol.com/ [accessed 2017-10-25] [WebCite Cache ID 6uU80peZq]
36. Ozawa K, Takemi Y, Eto K, Tanaka H, Fujii H, Ishikawa M, et al. Stage of vegetable intake and self-reported number of vegetable dishes consumed: are they valid measures of vegetable intake among middle-aged adults? [in Japanese]. Jpn J Nutr Dietetics 2013;71(3):97-111. [doi: 10.5264/eiyogakuzashi.71.97]
37. Yoshiike N, Hayashi F, Takemi Y, Mizoguchi K, Seino F. A new food guide in Japan: the Japanese food guide Spinning Top. Nutr Rev 2007 Apr;65(4):149-154. [Medline: 17503709]
38. Ministry of Health, Labour and Welfare. The National Health and Nutrition Survey [in Japanese]. 2009. URL: http://www.mhlw.go.jp/bunya/kenkou/eiyou/dl/h21-houkoku-01.pdf [accessed 2017-10-19] [WebCite Cache ID 6uKbYbIwS]
39. Bureau of Social Welfare Public Health. 2013. Tokyo Metropolitan Health Promotion Plan 21 (second): awareness and activities survey on ties with health and local residents [in Japanese] URL: http://www.fukushihoken.metro.tokyo.jp/kensui/plan21/chousa.html [accessed 2017-10-19] [WebCite Cache ID 6uKbdxE8J]
40. Blanchard CM, Fisher J, Sparling PB, Shanks TH, Nehl E, Rhodes RE, et al. Understanding adherence to 5 servings of fruits and vegetables per day: a theory of planned behavior perspective. J Nutr Educ Behav 2009 Jan;41(1):3-10. [doi: 10.1016/j.jneb.2007.12.006] [Medline: 19161914]
41. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869. [Medline: 20332511]
42. Ball K, McNaughton SA, Le HN, Abbott G, Stephens LD, Crawford DA. ShopSmart 4 Health: results of a randomized controlled trial of a behavioral intervention promoting fruit and vegetable consumption among socioeconomically disadvantaged women. Am J Clin Nutr 2016 Aug;104(2):436-445. [doi: 10.3945/ajcn.116.133173] [Medline: 27413129]
43. Kushida O, Murayama N. Effects of environmental intervention in workplace cafeterias on vegetable consumption by male workers. J Nutr Educ Behav 2014 Sep;46(5):350-358. [doi: 10.1016/j.jneb.2014.05.001] [Medline: 24974354]
44. Sawada K, Takemi Y, Murayama N, Sasaki S, Ishida H. Changes in vegetable consumption following intervention of a food environment intervention program in the employee cafeteria [in Japanese]. Jpn J Nutr Dietetics 2013;71(5):253-263. [doi: 10.5264/eiyogakuzashi.71.253]
45. Kristal AR, Glanz K, Curry SJ, Patterson RE. How can stages of change be best used in dietary interventions? J Am Diet Assoc 1999 Jun;99(6):679-684. [doi: 10.1016/S0002-8223(99)00165-0] [Medline: 10361529]
46. McDermott MS, Oliver M, Simnadis T, Beck EJ, Coltman T, Iverson D, et al. The Theory of Planned Behaviour and dietary patterns: a systematic review and meta-analysis. Prev Med 2015 Dec;81:150-156. [doi: 10.1016/j.ypmed.2015.08.020] [Medline: 26348455]
47. Beresford SA, Thompson B, Feng Z, Christianson A, McLerran D, Patrick DL. Seattle 5 a Day worksite program to increase fruit and vegetable consumption. Prev Med 2001 Mar;32(3):230-238. [doi: 10.1006/pmed.2000.0806] [Medline: 11277680]
48. Imanaka M, Ando M, Kitamura T, Kawamura T. Effectiveness of web-based self-disclosure peer-to-peer support for weight loss: randomized controlled trial. J Med Internet Res 2013 Jul;15(7):e136 [FREE Full text] [doi: 10.2196/jmir.2405] [Medline: 23838533]
49. Williams LK, Thornton L, Ball K, Crawford D. Is the objective food environment associated with perceptions of the food environment? Public Health Nutr 2012 Feb;15(2):291-298. [doi: 10.1017/S1368980011001947] [Medline: 21835079]
50. Statistics Bureau, Ministry of Internal Affairs and Communications. 2010. Population census: summary of the results and statistical tables URL: http://www.stat.go.jp/english/data/kokusei/2010/summary.htm [accessed 2017-10-19] [WebCite Cache ID 6uKcUNqut]

Abbreviations

RCT: randomized controlled trial
SES: socioeconomic status
TTM: transtheoretical model

Edited by G Eysenbach; submitted 14.05.17; peer-reviewed by J Lieffers, K Hicks; comments to author 13.07.17; revised version received 16.08.17; accepted 23.09.17; published 24.11.17.

Please cite as:
Nakamura S, Inayama T, Harada K, Arao T
Reduction in Vegetable Intake Disparities With a Web-Based Nutrition Education Intervention Among Lower-Income Adults in Japan: Randomized Controlled Trial
J Med Internet Res 2017;19(11):e377
URL: http://www.jmir.org/2017/11/e377/
doi:10.2196/jmir.8031
PMID:29175810

Original Paper

Implementation of a Web-Based Organ Donation Educational Intervention: Development and Use of a Refined Process Evaluation Model

Nakeva Redmond[1], MPH; Laura Harker[2], MPH; Yvan Bamps[3], PhD; Shauna St. Clair Flemming[1], MPH, MDiv; Jennie P Perryman[4], RN, PhD; Nancy J Thompson[1], MPH, PhD; Rachel E Patzer[5,6], MPH, PhD; Nancy S DeSousa Williams[1], MPH, PhD; Kimberly R Jacob Arriola[1], MPH, PhD

[1]Department of Behavioral Sciences and Health Education, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[2]Georgia Budget and Policy Institute, Atlanta, GA, United States

[3]Department of Neurosurgery, Emory University School of Medicine, Emory University, Atlanta, GA, United States

[4]Emory Transplant Center, Emory University Hospital, Emory Healthcare, Atlanta, GA, United States

[5]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[6]Department of Surgery, Emory University School of Medicine, Emory University, Atlanta, GA, United States

Corresponding Author:
Nakeva Redmond, MPH
Department of Behavioral Sciences and Health Education
Rollins School of Public Health
Emory University
1518 Clifton Road NE
GCR 426
Atlanta, GA, 30322
United States
Phone: 1 4047272386
Email: nakevar@gmail.com

Abstract

Background: The lack of available organs is often considered to be the single greatest problem in transplantation today. Internet use is at an all-time high, creating an opportunity to increase public commitment to organ donation through the broad reach of Web-based behavioral interventions. Implementing Internet interventions, however, presents challenges including preventing fraudulent respondents and ensuring intervention uptake. Although Web-based organ donation interventions have increased in recent years, process evaluation models appropriate for Web-based interventions are lacking.

Objective: The aim of this study was to describe a refined process evaluation model adapted for Web-based settings and used to assess the implementation of a Web-based intervention aimed to increase organ donation among African Americans.

Methods: We used a randomized pretest-posttest control design to assess the effectiveness of the intervention website that addressed barriers to organ donation through corresponding videos. Eligible participants were African American adult residents of Georgia who were not registered on the state donor registry. Drawing from previously developed process evaluation constructs, we adapted reach (the extent to which individuals were found eligible, and participated in the study), recruitment (online recruitment mechanism), dose received (intervention uptake), and context (how the Web-based setting influenced study implementation) for Internet settings and used the adapted model to assess the implementation of our Web-based intervention.

Results: With regard to reach, 1415 individuals completed the eligibility screener; 948 (67.00%) were determined eligible, of whom 918 (96.8%) completed the study. After eliminating duplicate entries (n=17), those who did not initiate the posttest (n=21) and those with an invalid ZIP code (n=108), 772 valid entries remained. Per the Internet protocol (IP) address analysis, only 23 of the 772 valid entries (3.0%) were within Georgia, and only 17 of those were considered unique entries and could be considered for analyses. With respect to recruitment, 517 of the 772 valid entries (67.0%) of participants were recruited from a Web recruiter. Regarding dose received, no videos from the intervention website were watched in their entirety, and the average viewing duration was 17 seconds over the minimum. With respect to context, context analysis provided us with valuable insights into factors in

the Internet environment that may have affected study implementation. Although only active for a brief period of time, the Craigslist website advertisement may have contributed the largest volume of fraudulent responses.

Conclusions: We determined fraud and low uptake to be serious threats to this study and further confirmed the importance of conducting a process evaluation to identify such threats. We suggest checking participants' IP addresses before study initiation, selecting software that allows for automatic duplicate protection, and tightening minimum requirements for intervention uptake. Further research is needed to understand how process evaluation models can be used to monitor implementation of Web-based studies.

(J Med Internet Res 2017;19(11):e396) doi:10.2196/jmir.8501

KEYWORDS

Internet; intervention; evaluation methodology; program evaluation; research techniques; organ donation; health education

Introduction

Using Web-Based Behavioral Interventions

The lack of available organs is often considered to be the single greatest problem in transplantation today [1]. As of July 2017, there were more than 117,000 persons awaiting a life-saving organ in the United States; however, less than 28,000 deceased donor transplants were performed in 2016 [2].

Efforts to address the growing gap between organ supply and demand include an extensive list of face-to-face educational interventions, with the majority of these interventions designed to improve attitudes and knowledge related to organ donation and promote positive donation intentions [3-11]. The widespread use of the Internet, however, creates an opportunity to increase public commitment to organ and tissue donation through innovative Web-based educational interventions. Approximately 88% of US adults currently use the Internet, with all demographic groups showing an exponential increase in use in recent years [12]. Internet-based interventions within the context of behavioral health research have capitalized on the high accessibility of the Internet and proliferated in the past 20 years [13], particularly in topics such as human immunodeficiency virus [14-16], depression and anxiety [17,18], and eating disorders [19,20], whereas Internet-based interventions addressing organ donation have only minimally been explored. Despite the pervasiveness of Web-based interventions, process evaluation frameworks able to assess their implementation are lacking.

Web-based interventions offer several advantages including fast and easy participant recruitment and data collection [21-27] and minimal cost compared with in-person interventions [15]. Internet interventions have been shown to have high fidelity, are transferrable to a variety of settings, and are scalable [15]. The added anonymity of the Internet reduces potential issues of social desirability and stigma, which is particularly appealing for studies addressing sensitive health topics. In addition to significantly improving knowledge, attitudes, and associated behaviors for various health topics, there is some evidence that compared with the in-person versions of the intervention, Web-based versions may be more effective [17,20,28]. Web-based organ donation initiatives have been successful in improving organ donation–related knowledge [29], attitudes [29,30], and behaviors [30], and reaching minority populations [31] who are consistently underrepresented as donors [2].

Assessing Fraud in Internet Studies

Despite these benefits, Internet studies present a list of disadvantages including high levels of attrition [32] and inability to achieve a representative sample [33]. Anonymity on the Internet can be considered a strength, but it also adds the potential of fraudulent and repeat responders [34]. Fraud in Internet studies has previously been cited as rare [35] and not compromising to the overall quality of data collected in Web-based studies [34]. However, it might be that investigators underestimate the extent of fraudulent responses and their impact on the internal validity of the associated studies. Subject fraud often involves elaborate adaptive strategies to counter detection measures in place and run virtually undetected even during data analysis. In one sexual behavioral risk Web-based study, nearly 11% of all responses were deemed ineligible because of fraud or repeated entry, with one respondent responsible for 6% of all responses [36]. Similarly, in a Web-based study surveying men who have sex with men, 11.6% of all entries were determined to be fraudulent and a result of changing responses on the study screener to gain entry or repeated entry into the study [37]. Further analysis demonstrated that including these fraudulent data would have greatly skewed study findings [37], underscoring the threat of fraudulent responders.

Few studies have offered explicit methods for preventing fraud from occurring in Web-based studies. Suggestions occur at differing stages of the research process and range from simple strategies such as eyeing suspicious entries (such as designated birthday does not match self-reported age) [35], asking participants whether they have already participated in the study before initiating [34,38], or flagging repeated emails and usernames [35,37-39], to more complex approaches such as altering study design (ie, requiring a telephone interview before participation) [35,36] or monitoring participants' Internet protocol (IP) addresses to determine participants' geographic locations and repeat responders [35]. Despite the range of options to prevent fraud, all methods have been found to be problematic [35,38] and sometimes labor intensive, with even the more advanced fraud detection methods requiring some level of manual checking [35,36]. The emerging use of "bots" or automated devices designed to hack studies without human effort, present an additional, more complex obstacle for preventing fraud and protecting study data. And although CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is used to prevent "bots" from accessing Internet forms (usually prompting users to type in the

name of an image or check a box), this function is not always successful in blocking these fraudulent attempts and may unintentionally deter valid participants with low computer literacy [35]. The ever-changing landscape of the Internet, and consequently Web-based interventions, requires equally dynamic security measures [35] and an understanding that confirming eligibility among all participants may not be possible in Web-based studies [36].

Assessing Intervention Uptake in Internet Studies

An additional concern in implementing Web-based studies is ensuring uptake of the intervention. Most Web-based study participants have the convenience of completing the study on devices in their own homes and on their own time. Whereas such autonomy within the context of a research study may alleviate accessibility issues among the target population [21], it also limits implementer control and presents the risk of insufficient uptake of the intervention. Overall, measured intervention uptake on Web-based platforms has been found to be less than intended [40], as participants have been found to abandon the intervention before finishing [40] and "clicking through" associated questionnaires without reading all of the content [41].

Ensuring uptake of Web-based interventions promoting organ donation is especially challenging. Organ donation registration has been characterized as a low-vested-interest health behavior that is accompanied by high attitudinal ambivalence and a lack of personal benefit [42], thus making it difficult to appeal to the general population. Moreover, there is evidence that distrust of the medical care system and physicians undergird negative attitudes about organ and tissue donation, particularly among African Americans [43]. These conditions may make uptake of a Web-based organ donation education intervention particularly challenging within this population.

Although Web-based studies are more difficult to control, Web-based analytics allow study implementers to unobtrusively monitor participants' engagement with the study [41], with log-in rates, website hits (number of people who landed on the website), pages visited, modules completed, and completion of the entire intervention serving as the most commonly used exposure measures [40]. Monitoring uptake is not only becoming more common but expected for Internet studies [21], creating a dual opportunity to assess participant engagement and potentially identify fraud [44] by monitoring patterns of website use.

Measuring uptake of in-person intervention materials is common practice and helps attribute specific changes in study outcomes to varying levels of exposure to the intervention [45]. Measuring uptake for Web-based studies, however, is less explored, and has yielded mixed results in terms of project outcomes. For example, Christensen et al [46] measured intervention uptake of a Web-based cognitive behavioral therapy study and found that participants' anxiety and depression scores significantly lowered with each additional module completed and more time spent on the modules. In a Web-based smokeless tobacco cessation program, Danaher and Seeley [21] found Web-based program exposure to mediate the relationship between self-efficacy and tobacco abstinence. And although it may be expected that increased exposure to intervention materials is positively associated with outcomes, Glasgow et al [32] found that ongoing engagement was negatively associated with the amount of material presented in the intervention, which of course presents a potentially new threat to the study design and to desired outcomes.

Evaluating Implementation of Internet-Based Interventions

Some efforts have been made to determine how to implement Internet interventions in ways that maximize their effectiveness. For example, Cummins et al (2003) created the 5 As (advise, assess, assist, anticipatory guidance, and arrange follow-up) for effective health behavior change treatment on the Internet framework to address an emerging need for evaluation frameworks for Internet studies [47]. However, this model is limiting as it was adapted for preventive behaviors and disease management mostly focused on assessing content and admittedly only "provided the minimum criteria for a program to have the potential for providing behavior change" [47]. More recently, O'Grady et al (2009) created a dynamic evaluation framework for interactive Web technologies [48]. Although this framework is extensive, it maintains a conceptual approach and leaves the question of how to practically implement such a robust evaluation model, which is an overall critique of Web-based evaluation models [49]. Additional evaluation efforts on the Internet have been designed for static websites that provide health information and not interactive websites [50] and have mostly relied on self-report to assess usability [51], quality [51], and uptake [21,40,52,53] of the website. Whereas these evaluative techniques are insightful, more rigorous and objective measures are necessary for implementers to fully explore implementation of interactive Web-based interventions. Process evaluation is necessary to explain results according to intervention component, assess quality, and accuracy [45] and to help explain why a program was or was not successful [54]. Conducting a process evaluation to understand how Internet studies are implemented is imperative to optimize the Web-based platform, ensure program uptake, minimize waste, and maximize intervention effectiveness.

The purpose of this study was to investigate how a culturally sensitive, Web-based intervention aimed to increase organ donation among African Americans in Georgia was implemented by using a process evaluation model adapted for Web-based interventions. Drawing from previously developed process evaluation constructs [45], we adapted reach, recruitment, dose received, and context for use in Internet-settings. We excluded dose delivered and implementation from our adapted model, as all intervention components were delivered on a standardized Web-based platform. We also did not formally include fidelity (an original construct measuring the overall quality of intervention implementation) in our process analyses because of its strong relationship with the included process evaluation constructs and inherent measurement challenges [45]. Relevant to each process evaluation construct, this study sought to answer six research questions:

1. Reach: What proportion of interested participants was found eligible for the study?

2. Reach: What proportion of eligible participants participated in the study?
3. Reach: To what extent were Georgia residents recruited into the study?
4. Recruitment: To what extent were Web recruiters successful in attracting participants to the study?
5. Dose received: To what extent did participants engage in the study intervention components?
6. Context: To what extent did the Web-based setting influence study implementation?

Methods

Study Design and Sample

Project WEB ACTS (About Choices in Transplantation and Sharing, ACTS) is a Web-based intervention designed to assess the feasibility of using the Internet to deliver culturally sensitive, educational materials about organ donation and to increase donor registration among African American adults in Georgia. We used a randomized controlled pretest-posttest design to compare the effectiveness of the intervention website (Project WEB ACTS) to the control website (Donate Life Georgia). The Donate Life Georgia website features general facts about organ and tissue donation and personal donation stories, both nonspecific to any particular racial group. There is also a link to sign up on the Georgia state donor registry on the Donate Life website. Eligible participants identified as African American or black, were residents of Georgia, were 18 years of age or older, and were not already registered on the Georgia state donor registry, as indicated by self-report. This study was reviewed by the Emory University Institutional Review Board and considered exempt from further review and approval (IRB00078995).

Website Design

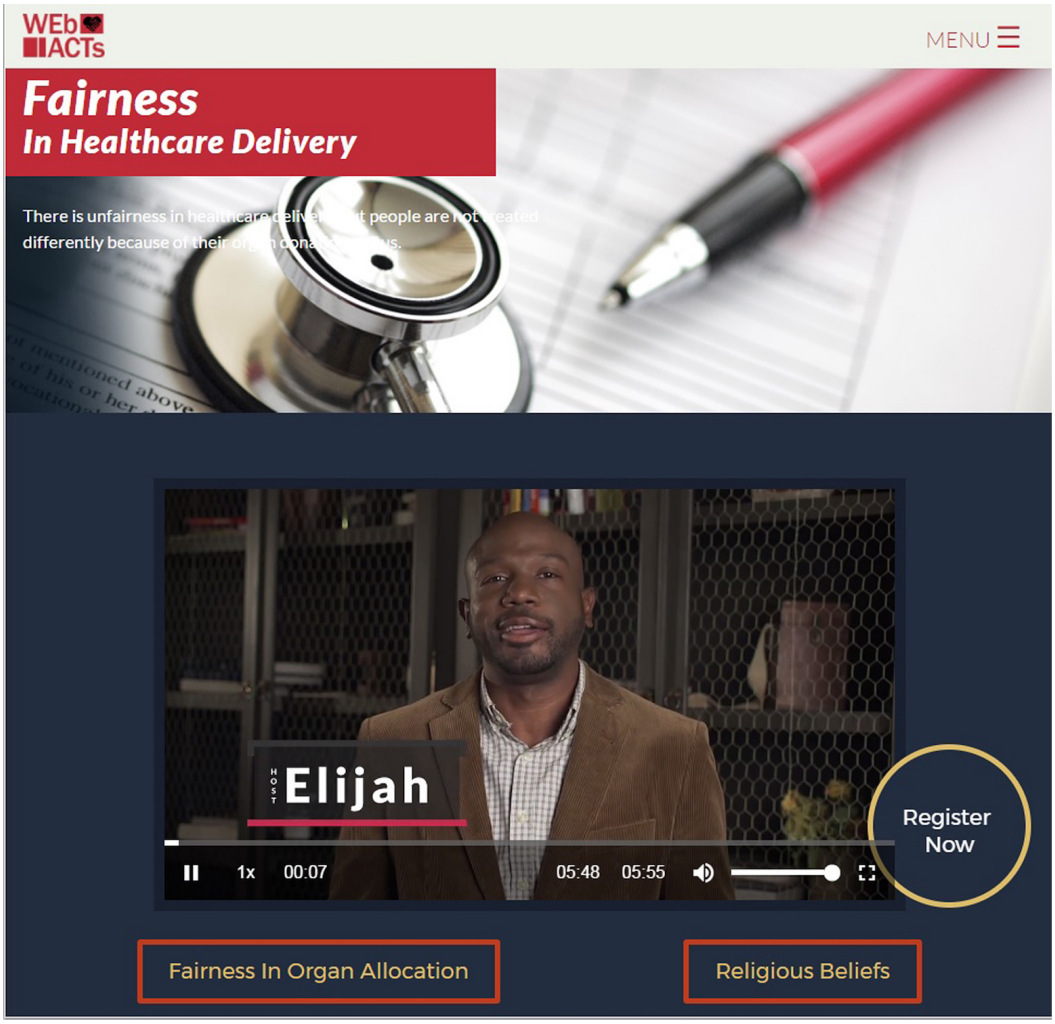
Using organ donation literature, the two-dimensional model of cultural sensitivity [55], and the IIFF model of donor registration [42], we identified five prominent organ donation topic areas relevant for African Americans [56,57]. The Project WEB ACTS website modules are Act Now, Fairness in Organ Allocation, Fairness in Health Care Delivery, Religious Beliefs, and Let's Talk About Life. Each module contains a corresponding 3- to 4-min video featuring an African American host and individuals discussing the importance of organ donation from various perspectives (ie, pastors, physicians, and living donors). The final tab of the website features a link leading to the Georgia state donor registry for those participants who decide to register as an organ donor at the conclusion of the intervention. A website consulting company was hired to develop the Project WEB ACTS website. All content was adapted from prior iterations of Project ACTS [58-60] and informed by both formative and scientific research (Figure 1).

Data Collection Procedures

We identified and trained 17 Web recruiters to recruit participants into the study, primarily using Web-based mechanisms (eg, social media, Web-based listservs, and electronic bulletin boards). Web recruiters were demographically similar to the target population, familiar with Internet technology, and maintained strong online networks. Using approved language, Web recruiters advertised the Project WEB ACTS study, with each posting including a brief description of the project, mention of the incentive (US $10 virtual gift card), and a URL link to the study eligibility screener. The participant incentive was considered appropriate for the anticipated time burden of the proposed intervention according to prior Web-based studies [32,36], yet was kept nominal to reduce the risk of repeat participants [39]. Convenience sampling was used to identify eligible participants. There was a strong emphasis on online recruitment; however, recruiters were also permitted to share paper recruiting materials, including hard copy flyers with their in-person social networks (eg, at church, work, or in their neighborhood). Additionally, Project WEB ACTS administrators briefly (for approximately 24 hours) advertised the study on the Craigslist website to recruit participants into the study.

The project coordinator determined individuals' eligibility based on answers to a Web-based eligibility screener and informed participants of their eligibility status via email. Emails to eligible participants included a link leading to the study consent and baseline assessment (a 38-item questionnaire measuring attitudes and beliefs regarding donation and transplantation, knowledge of the donation and transplantation system, Internet usage, and demographic characteristics). After completing the baseline assessment, participants were randomized to either the intervention or control website and required to spend a minimum of 10 min on the site before the posttest questionnaire was made available. The posttest questionnaire consisted of 24 items, including the attitudes, beliefs, and knowledge measures from the pretest and an added two items measuring the likelihood that the respondent would register as an organ donor in the future. Both intervention and control websites included a "call to action" for donation registration, and participants were allowed to sign up as organ donors on the Georgia state donor registry while on either website. Control and intervention participants were provided with a code upon completion of the posttest questionnaire and instructed to email the project coordinator with the code to receive the incentive. Each email address was checked before distributing the virtual gift card to ensure that multiple incentives were not sent to the same email address.

Figure 1. Intervention (Project WEB ACTS) interactive website, Fairness In Healthcare Delivery module.



Adaptation and Implementation of Process Measures

We first analyzed frequencies of all usable data using Statistical Package for the Social Sciences (SPSS) version 23.0 (IBM Corp). Drawing from previously developed process evaluation constructs [45,54,61], we adapted reach, recruitment, dose received, and context for Internet-based interventions and used each to conduct a process evaluation of our study (Table 1).

Reach

Reach is most commonly considered the proportion of the target audience that participates in the study [45,54]. We can estimate our target audience (African American adults living in Georgia and not registered as organ donors) given census data and state registry data; however, those individuals who visited the study eligibility site better represent our target, given the absence of any widespread, population-level marketing. The breadth of the Internet and our use of convenience sampling through online networks make it difficult to determine a numeric value of the intended target audience, requiring the adaptation of the reach construct. As a result, we defined reach as the proportion of interested participants who were found eligible for the study and the proportion of eligible participants who participated in the study. The study's project coordinator manually checked each participant's eligibility screener before study entry.

We also defined reach as the extent to which Georgia residents were recruited into the study. Although the Project WEB ACTS study was administered entirely on the Web, it was important to assess the physical environment of participants, as this study sought to explicitly recruit Georgia residents to register for the Georgia state donor registry. States maintain separate donor registries; therefore, including participants outside of the state of Georgia would undermine our ability to document registration on the Georgia state donor registry. We used IP addresses to approximate each participant's location at the time of the study and to assess the extent to which Georgia residents were recruited into the study. First, we retrospectively analyzed each participant's IP address to determine his or her approximate location at the time of completing the study and to identify repeated IP addresses. Similar to a street address, an IP address is a unique set of numbers used to identify computers on a network [39]. Whereas multiple individuals who share a home or work in the same building may be able to use one IP address, we expected such instances to be minimal. We used a reputable Internet IP address search engine to obtain IP information and used a second recommended IP address search engine to spot check results from the first search engine. IP addresses were used to estimate participant location (United States vs non-United States and US state) and to identify study repeaters.

Table 1. Adapted process measures applied to Project WEB ACTS (About Choices in Transplantation and Sharing) intervention.

Construct	Definition	Research question	Indicators
Reach	The extent to which individuals are found eligible, and participate in the study	What proportion of interested participants was eligible for the study?	Percentage of individuals who completed the eligibility questionnaire and were found eligible for the study
		What proportion of eligible participants participated in the study?	Percentage of individuals who were found eligible per the eligibility questionnaire and completed the study
		To what extent were Georgia residents recruited into the study?	Location estimation (Georgia vs other US state) per retrospective Internet protocol address analysis
Recruitment	Procedures used to approach and attract participants	To what extent were Web recruiters successful in attracting participants to the study?	Percentage of participants referred to study from Web recruiters and other methods
Dose received	The extent to which participants actively engage with, interact with, are receptive to, or use materials or recommended resources	To what extent did participants engage in the study intervention components?	• Number of website sessions • Percentage of new sessions • Number of users • Average time spent on the entire website (min:sec) • Number of page views • Average time on each page (min:sec) • Number of individuals who pressed "play" for each video • Highest percentage completed of each video • Number of individuals completing the corresponding highest percent completed
Context	Aspects of the Web-based setting that may influence intervention implementation	To what extent did the Web-based setting influence intervention implementation?	• Data trends of eligibility screener responses as compared with Web-based marketing efforts • Debriefing interviews with Web recruiters

Respondents with IP addresses outside of the United States, IP addresses outside of the state of Georgia, and chronically repeating IP addresses (defined as those repeating more than 3 times) were identified as having concerning IP addresses. All concerning entries were flagged for further consideration.

Recruitment

Participants were asked to designate how they were recruited into the study, with answer choices including Web recruiters (listed by name) or other methods (which we inferred to be our Craigslist website advertisement). Given the emphasis on recruitment through social media and Internet forums and the difficulty in objectively tracking recruitment, recruitment source was measured by self-report.

Dose Received

We used Google analytics software to assess dose received or the extent to which participants were exposed to the WEB ACTS intervention website. This software allowed us to collect aggregate data on uptake of the website, including number of users, number of page views, average time spent on each page, and highest percentage completed of each of the five videos. Participant usage data were not available for control participants accessing the Donate Life Georgia website because it is a public webpage.

Context

Context in traditional study settings considers the "larger physical, social, and political environment that either directly or indirectly affects an intervention" [45]. The Internet is its own environment, with its own culture [44]; thus, context may be assessed as the Web-based setting to determine what factors in the virtual study environment may have affected intervention implementation. The dynamic nature of the Internet, however, makes it difficult to monitor activity broadly. We focused on Web-based environments where the study was advertised, including social media pages and Craigslist website, with the understanding that assessing where the study link could be accessed could provide information regarding most effective recruitment sites, sites potentially encouraging fraud, and other contextual factors affecting overall study implementation. We used data trends as a proxy indicator for assessing context, including responses to the eligibility screener as compared with Web-based marketing efforts and also conducted debriefing interviews with Web recruiters to gain insights to events happening on social media sites used for recruitment and to corroborate our data.

Results

Participant Demographics

Project WEB ACTS opened for recruitment on September 21, 2015 and closed on November 13, 2015, and a total of 772 participants were included in our initial participant pool. The majority of participants were male (56.2%, 434/772), and ages ranged from 18 to 74 years with a median age of 37 years. Most of the participants were married (74.8%, 577/772). Approximately half of all the participants reported college as the highest level of education completed, and the vast majority of participants were working full-time or part-time (94.3%, 728/772; Table 2).

Reach

A total of 1415 individuals completed the eligibility screener; 948 (67.00%, 948/1415) were determined to be eligible based upon their responses to this screener. Of the 467 (33.00%, 467/1415) ineligible participants, 109 were not African American, 6 were not residents of Georgia, and 449 indicated that they were already on the Georgia state donor registry (criteria were not mutually exclusive). Of the 948 eligible participants, 918 (96.8%) completed the study. However, after eliminating duplicate entries (those with repeat emails or usernames; n=17), those who did not initiate the posttest (n=21) and those who provided an invalid ZIP code (n=108), we ended with a total of 772 surveys to use for further analysis.

Per the IP address poststudy analysis, 600 (77.7%, 600/772) respondents had an IP address within the United States, whereas 172 (22. 3%, 172/772) had an IP address outside of the United States. Of those in the United States, 577 (96.2%, 577/600) had IP addresses outside of Georgia, and 23 (3.8%) had IP addresses within Georgia. Of the 23 respondents who completed the study within Georgia, 17 (74%, 17/23) entries were considered unique entries. In sum, only 2.2% (17/772) of our original 772 surveys included for analysis were deemed usable data at the conclusion of our analyses.

Although respondents with IP addresses outside of Georgia (United States and international) were not considered usable data, all IP addresses were analyzed to determine the extent of the threat of repeat responders to the entire study. Of the 577 international IP addresses, 207 (35.9%) completed the study more than one time. Of the 600 IP addresses within the United States, 144 (24.0%) completed the study more than one time. Chronic repeaters were seen among both US and international IP addresses, with respondents with international IPs presenting a more aggressive threat. Respondents with international IP addresses repeated up to 47 times, with the next highest repeats totaling 35 and 36 times. Respondents with US IPs repeated up to 35 times, with the next highest repeat being 18 times from one IP address (Figure 2; data points are not mutually exclusive).

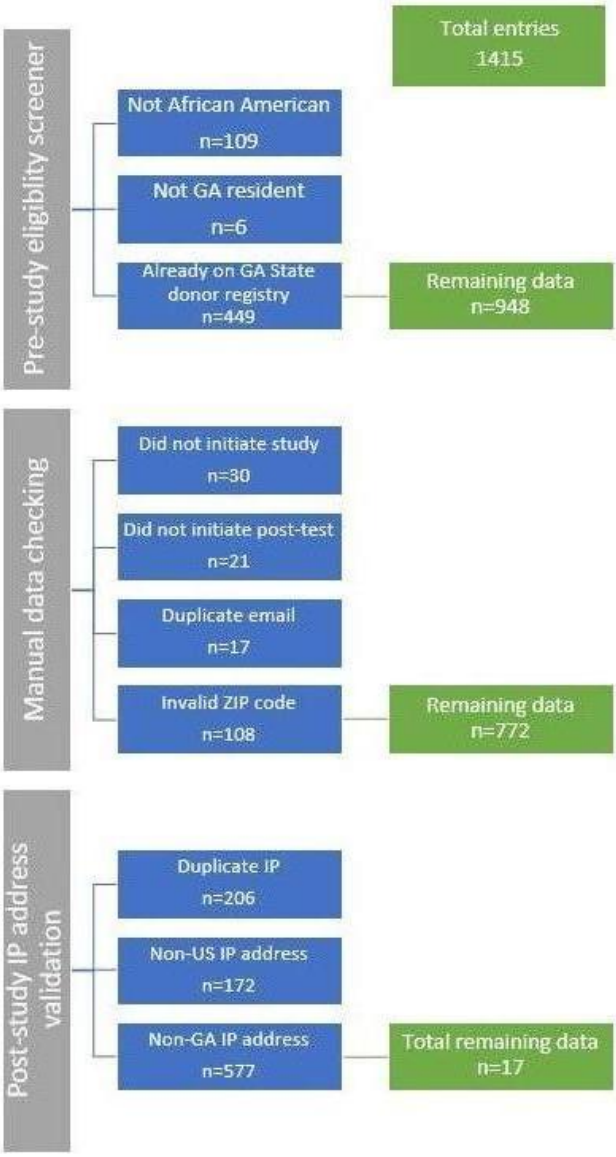
Recruitment

Approximately 517 participants of the 772 entries included after manual data cleaning (67.0%) reported that they were directly recruited from a trained Web recruiter, and the remainder from other means (eg, Craigslist website posting). Each Web recruiter recruited between 11 and 62 participants, with an average of 32 participants per recruiter.

Table 2. Participant demographics (N=772; sample sizes and percentages vary because of missing data).

Characteristic	Frequency
Age, mean (range)	36.16 (18-74)
Gender, n (%)	
Female	332 (43.0)
Male	434 (56.2)
Education, n (%)	
Less than high school	17 (2.2)
12th grade or GED	292 (37.8)
College	382 (49.5)
Professional degree	71 (9.2)
Employment, n (%)	
Unemployed	15 (1.9)
Retired	12 (1.6)
Working part-time or full-time	728 (94.3)
Marital status, n (%)	
Single	158 (20.5)
Married	577 (74.7)
Divorced or separated or widowed	17 (2.2)

Figure 2. Summary of eliminated data through Reach assessment.



Dose Received

Website usage analytics showed that 1056 individuals accessed the study website home page (including intervention and control participants). Approximately 90.13% of these website visits were considered "new" or accessed by different individuals. A total of 417 participants were randomized to the intervention group; only two entries of the intervention group were among the 17 data entries deemed as valid from our reach analysis. Those in the intervention group who viewed the WEB ACTS website had an average session duration of 10 min 17 s, just 17 s over the required minimum. Activity on each of the WEB ACTS modules decreased as participants moved to the next module. The first module (Act Now), for example, had 538 page views (1.29 page views per user), and the average time spent on the page was about 6 min. The shift from the first module to the next module, Fairness in Organ Allocation, produced the greatefst drop-off in participation, with only 149 page views (0.36 page views per user) and an average page time of about 2 min. Participation continued to lessen with each subsequent module, with the final module, Let's Talk About life, having only 68 page views (0.16 page views per user) but a slight peak in average viewing time (2:20 average page time).

Video viewing for each module was also minimal. Thirty-four participants pressed play for the Act Now video, 10 for the Fairness in Organ Allocation video, and 8 for the Fairness in Health Care Delivery video. The most of any video watched was the Fairness in Organ Allocation video; one person watched 70% of the video. No videos were watched in their entirety (Table 3).

Context

Minimal study activity was observed during the first weeks the study was active. Responses to the eligibility screener averaged about 1.61 responses per day. Response to the study increased significantly once the Craigslist ad was placed; a total of 485 responses were recorded during the 24 hours that the ad was posted, and responses averaged 55.12 per day from the time the advertisement was removed until the study closed.

Table 3. WEB ACTS (About Choices in Transplantation and Sharing) website usage statistics (n=417).

Module[a]	Title	Number of page views (views per person)	Average time on page (min:sec)	Number of times pressed "play" for video	Highest percentage of video completed n (%)
Module I	Act Now	538 (1.29)	06:02	34	2 (50)
Module II	Fairness in Organ Allocation	149 (0.36)	02:08	10	1 (70)
Module III	Fairness in Health Care Delivery	108 (0.26)	01:28	8	1 (20)
Module IV	Religious Beliefs	104 (0.25)	01:25	11	1 (10)
Module V	Let's Talk About Life	68 (0.16)	02:20	13	1 (50)

[a]Module videos were of different lengths and as follows: module I: 02:40, module II: 05:15, module III: 05:55, module IV: 05:43, and module V: 04:02.

The proportion of respondents found eligible increased the longer the study stayed active and particularly shifted at a certain time point of the study. Nearly 88.9% (431/485) of respondents on the first to the 28th day of the study were deemed ineligible; on the 29th day to day 60 (the end of the study), only 10.97% (102/930) of respondents were determined to be ineligible.

Web recruiters noted mostly using Facebook to recruit participants and said that they posted about the study on their personal accounts 3 to 4 times per day, on average. Some Web recruiters mentioned their Facebook friends reposting the study on their own pages, although we are not sure to what extent this occurred during the study. Overall, Web recruiters said they received minimal feedback about the study from friends on social media, but some mentioned being questioned whether the study was "real" or just spam.

Discussion

Principal Findings

This study sought to explore threats to Web-based interventions through the use of a process evaluation model refined for Internet settings. Whereas previous research suggests that fraud in Internet studies is minimal [35], we found fraudulent responses to be pervasive in our study, greatly impacting the overall quality of our data. Our study found that nearly 97.8% (755/772) of participants' entries were invalid, despite having provided eligible responses on the eligibility screener. This finding supports the notion that ineligible and repeat participants are eager to infiltrate Web-based studies to receive monetary incentives, potentially creating a high volume of invalid responses. With a range of security parameters for Web-based studies available and no clear protocol on which strategies to use or to what extent to use them, assessing reach following our adapted process measures helped inform how to better secure our data collection protocol. In addition to using a CAPTCHA function to potentially protect the study from "bots," we also learned that screening for IP addresses before study initiation is necessary to ensure study eligibility, particularly for studies requiring that participants reside in a particular geographical region.

Dose received analysis confirmed that uptake of the Web-based intervention was nominal, reaffirming findings regarding minimal uptake in prior studies [40,41]. Overall, website navigation for the intervention group was low, particularly as modules progressed. Additionally, no videos were watched in their entirety, and only 8% of participants pressed play for the first video, which was the most played of any video. Our initial study protocol implemented a minimal time requirement to remain on the website but did not mandate a minimal viewing time for website modules in the WEB ACTS site. This strategy was meant to encourage participants to explore modules in which they were most interested and that best fit their individually identified barriers to organ donation. And although we know that more uptake does not necessarily translate to more positive outcomes [32], given the negligible amount of uptake observed in our study, a minimum uptake requirement should be instituted to attribute any observed changes in outcomes to the intervention. It is important to note that the high volume of fraudulent respondents also potentially influenced minimal uptake of the intervention (ie, we would not expect video viewing to be high among fraudulent and repeat respondents who are seemingly accessing the study to receive multiple incentives). This interrelatedness of fraud and minimal uptake further underscores the importance of protecting Web-based studies from such threats.

Determining the optimal amount of study uptake to produce desired outcomes and prevent study attrition is difficult, as previous findings on what works varies [21,32,46], and to our knowledge, there are no published studies that outline how to determine this minimum. We suggest conducting efficacy testing before launching Web-based interventions to determine this optimal level of study consumption for specific study topics and demographics. In those cases where a mandate cannot be made because of technological limitations, knowing how much of an intervention individual participants consumed can be just as valuable, as these data would provide an opportunity to link consumption levels to outcomes [46]. Although we were only able to collect aggregate-level data to assess dose received, looking at individual consumption should be explored in future studies.

Context analysis provided us with valuable insights into factors in the Internet environment that may have affected study implementation. Although only active for a brief period of time, the Craigslist advertisement may have contributed the largest volume of fraudulent responses (our recruitment analysis determined that only 33.0% (255/772) of participants said they were recruited by the Craigslist ad; however, given the extent of fraudulent responses, these data were not considered to be reliable). We saw an exponential increase in responses to the

eligibility screener after the Cragislist advertisement was posted, and most of those responses were determined to be fraudulent. We also observed that as time passed, more individuals were determined to be eligible for the study, suggesting that these fraudulent responders may have learned the desired eligibility and augmented answers on the screener to gain access to the study.

Conducting an assessment of context provided valuable information pertaining to program implementation, but such an assessment was difficult to conduct in an Internet environment, partially because of limited access to data on outside websites used to advertise our study (such as IP addresses of individuals accessing Craigslist or Facebook clicks). Despite these challenges, we found that using proxy data as an indicator to assess context provided valuable insights regarding how the Web-based study environment may have affected implementation and suggest using similar indicators when assessing context in Web-based studies. Steckler and Linnan [45] advise outlining potentially influential contextual factors before study initiation. For example, access to local recreational facilities would be an important contextual factor to consider if implementing a physical activity intervention, as it may be that participants with greater access to such facilities experience greater success with the intervention [45]. We also anticipate this practice would be valuable in Web-based study environments and may involve monitoring activity on websites associated with study marketing before implementation and throughout the entire study process, particularly when using social media sites that may involve interactions between individuals.

Although our context analysis provided some explanation as to how our study design may have facilitated fraudulent responses, the question of why so few legitimate participants were recruited remains. We learned from debriefing interviews with Web-based recruiters that even though some potential participants learned of the study from Web recruiters who they knew personally, there was still some skepticism around participating in the study. This suspicion may stem from concerns about compromising Internet privacy by clicking on foreign links, and/or interpreting unfamiliar links as “spam” [62], or may also stem from the topic of organ donation itself. As previously stated, organ donation is an altruistic health behavior that is difficult to change [42] and may elicit skepticism from African Americans, in particular, given overall feelings of distrust toward the American health care and organ allocation system [43]. And whereas research has found the Internet to be a feasible modality for implementing behavioral studies on sensitive topics [14-16,19,20], introducing organ donation on the Internet may heighten uncertainty around a controversial topic. More research is needed to determine how Web recruiters can be used to recruit participants into Web-based research studies using social media (particularly for sensitive topics) and how to make organ donation research participation more appealing to the general public.

Limitations

There are several limitations to this study. First, portions of our process evaluation model were implemented retrospectively and not in tandem with study implementation, which would have been ideal to protect study data. For example, data regarding fraudulent participants through our reach analysis would have prompted us to implement tighter protocols mid-study, which could have led to much different overall results of the study. Second, some indicators used to measure our adapted process measures (such as data trends for context) were proxy measures and not direct measures of study implementation. And, whereas process evaluation indicators often rely on second-source feedback (such as interviews) [45], more vetting of these methods is needed to determine best practices for collecting this type of data via the Internet. Furthermore, because Web recruiter debriefing interviews were not required, only those recruiters who were interested provided feedback, allowing for some bias in the data collected. Finally, our adapted process evaluation model was only implemented on our study addressing organ donation behaviors among African Americans; more research is needed to see how this model can be used with Web-based interventions addressing other health topics among different populations.

Conclusions

We found our process evaluation of the WEB ACTS study to be helpful in assessing the implementation of a culturally sensitive, Web-based educational intervention. Fraudulent responses and insufficient intervention uptake were identified as threats to our study. On the basis of our findings, we suggest checking participants’ IP addresses before study initiation, selecting software that allows for automatic duplicate protection, and tightening minimum requirements for intervention uptake. Whereas these parameters may have some trade-offs (including potentially deterring some legitimate participants), we anticipate these threats to be minimal and protecting the study to be of the highest priority. Clearly, implementing Web-based studies is complex and begs for continual monitoring through process evaluation models fit for Web-based settings. Further research is needed to understand how process evaluation models can be used to monitor implementation of Web-based studies.

Acknowledgments

This work was supported by a grant from the Health Resources and Services Administration (R390T26991).

Authors' Contributions

NR, LH, JPP, NJT, REP, and KRJA are responsible for the design of the study. NR performed the analysis and wrote the first draft of the manuscript. KRJA contributed to the manuscript development. YB contributed conceptual and technical knowledge, which informed the manuscript. SSF, NJT, REP, NSDW, and KRJA revised the manuscript critically. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Ganikos M. Organ donation: an overview of the field. In: Siegel JT, Alvaro EM, editors. Understanding Organ Donation: Applied Behavioral Science Perspectives. Hoboken, NJ: Wiley-Blackwell; Jan 06, 2010:13-39.
2. U.S. Department of Health and Human Services. Optn.transplant.hrsa. 2015. Organ procurement and transplantation network URL: https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/# [accessed 2017-11-07] [WebCite Cache ID 6unrr5uxP]
3. Anantachoti P, Gross CR, Gunderson S. Promoting organ donation among high school students: an educational intervention. Prog Transplant 2001 Sep;11(3):201-207. [Medline: 11949463]
4. Birkimer JC, Barbee AP, Francis ML, Berry MM, Deuser PS, Pope JR. Effects of refutational messages, thought provocation, and decision deadlines on signing to donate organs. J Appl Social Pyschol 1994 Oct;24(19):1735-1761. [doi: 10.1111/j.1559-1816.1994.tb01572.x]
5. Cantarovich F, Fagundes E, Biolcalti D, Bacqué MC. School education, a basis for positive attitudes toward organ donation. Transplant Proc 2000 Feb;32(1):55-56. [Medline: 10700966]
6. Carducci BJ, Denser PS, Bauer A, Large M, Ramaekers M. An application of the foot in the door technique to organ donation. J Bus Psychol 1989 Dec;4(2):245-249. [doi: 10.1007/BF01016444]
7. Kopfman JE, Smith SW. Understanding the audiences of a health communication campaign: a discriminant analysis of potential organ donors based on intent to donate. J Appl Comm Res 2009 May 21;24(1):33-49. [doi: 10.1080/00909889609365438]
8. Kopfman JE, Smith SW, Ah Yun JK, Hodges A. Affective and cognitive reactions to narrative versus statistical evidence organ donation messages. J Appl Commun Res 1998;26(3):279-300. [doi: 10.1080/00909889809365508]
9. Radecki CM, Jaccard J. Psychological aspects of organ donation: a critical review and synthesis of individual and next-of-kin donation decisions. Health Psychol 1997 Mar;16(2):183-195. [Medline: 9269891]
10. Smith SW, Morrison K, Kopfman JE, Ford LA. The influence of prior thought and intent on the memorability and persuasiveness of organ donation message strategies. Health Commun 1994 Jan 01;6(1):1-20. [doi: 10.1207/s15327027hc0601_1]
11. Weaver M, Spigner C, Pineda M, Rabun KG, Allen MD. Knowledge and opinions about organ donation among urban high school students: pilot test of a health education program. Clin Transplant 2000 Aug;14(4 Pt 1):292-303. [Medline: 10945199]
12. Pew Research Center. Pewinternet. 2017 Jan 11. Internet use URL: http://www.pewinternet.org/chart/internet-use/ [accessed 2017-11-08] [WebCite Cache ID 6upM3xJpI]
13. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 2008 Jul;47(1):3-16 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]
14. Castillo-Arcos Ldel C, Benavides-Torres RA, López-Rosales F, Onofre-Rodríguez DJ, Valdez-Montero C, Maas-Góngora L. The effect of an Internet-based intervention designed to reduce HIV/AIDS sexual risk among Mexican adolescents. AIDS Care 2016;28(2):191-196. [doi: 10.1080/09540121.2015.1073663] [Medline: 26301864]
15. Marsch LA, Guarino H, Grabinski MJ, Syckes C, Dillingham ET, Xie H, et al. Comparative effectiveness of web-based vs. educator-delivered HIV prevention for adolescent substance users: a randomized, controlled trial. J Subst Abuse Treat 2015 Dec;59:30-37 [FREE Full text] [doi: 10.1016/j.jsat.2015.07.003] [Medline: 26293644]
16. Jones R, Hoover DR, Lacroix LJ. A randomized controlled trial of soap opera videos streamed to smartphones to reduce risk of sexually transmitted human immunodeficiency virus (HIV) in young urban African American women. Nurs Outlook 2013;61(4):205-215.e3 [FREE Full text] [doi: 10.1016/j.outlook.2013.03.006] [Medline: 23743482]
17. O'Kearney R, Kang K, Christensen H, Griffiths K. A controlled trial of a school-based Internet program for reducing depressive symptoms in adolescent girls. Depress Anxiety 2009;26(1):65-72. [doi: 10.1002/da.20507] [Medline: 18828141]
18. Thompson NJ, Patel AH, Selwa LM, Stoll SC, Begley CE, Johnson EK, et al. Expanding the efficacy of Project UPLIFT: Distance delivery of mindfulness-based depression prevention to people with epilepsy. J Consult Clin Psychol 2015 Apr;83(2):304-313 [FREE Full text] [doi: 10.1037/a0038404] [Medline: 25495361]
19. Jones M, Taylor Lynch K, Kass AE, Burrows A, Williams J, Wilfley DE, et al. Healthy weight regulation and eating disorder prevention in high school students: a universal and targeted Web-based intervention. J Med Internet Res 2014 Feb 27;16(2):e57 [FREE Full text] [doi: 10.2196/jmir.2995] [Medline: 24583683]
20. Heinicke BE, Paxton SJ, McLean SA, Wertheim EH. Internet-delivered targeted group intervention for body dissatisfaction and disordered eating in adolescent girls: a randomized controlled trial. J Abnorm Child Psychol 2007 Jun;35(3):379-391. [doi: 10.1007/s10802-006-9097-9] [Medline: 17243014]
21. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]

22. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
23. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]
24. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
25. Lohse B. Facebook is an effective strategy to recruit low-income women to online nutrition education. J Nutr Educ Behav 2013;45(1):69-76. [doi: 10.1016/j.jneb.2012.06.006] [Medline: 23305805]
26. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. Nicotine Tob Res 2006 Dec;8 Suppl 1:S35-S41. [Medline: 17491169]
27. Watson B, Robinson DH, Harker L, Arriola KR. The inclusion of African-American study participants in web-based research studies: viewpoint. J Med Internet Res 2016 Jun;18(6):e168 [FREE Full text] [doi: 10.2196/jmir.5486] [Medline: 27334683]
28. Williamson DA, Martin PD, White MA, Newton R, Walden H, York-Crowe E, et al. Efficacy of an internet-based behavioral weight loss program for overweight adolescent African-American girls. Eat Weight Disord 2005 Sep;10(3):193-203. [Medline: 16277142]
29. Vinokur AD, Merion RM, Couper MP, Jones EG, Dong Y. Educational web-based intervention for high school students to increase knowledge and promote positive attitudes toward organ donation. Health Educ Behav 2006 Dec;33(6):773-786. [doi: 10.1177/1090198106288596] [Medline: 16923836]
30. Merion RM, Vinokur AD, Couper MP, Jones EG, Dong Y, Wimsatt M, et al. Internet-based intervention to promote organ donor registry participation and family notification. Transplantation 2003 Apr 27;75(8):1175-1179. [doi: 10.1097/01.TP.0000062845.51042.DD] [Medline: 12717199]
31. Lockwood MB, Saunders MR, Lee CS, Becker YT, Josephson MA, Chon WJ. Kidney transplant and the digital divide: is information and communication technology a barrier or a bridge to transplant for African Americans? Prog Transplant 2013 Dec;23(4):302-309. [doi: 10.7182/pit2013869] [Medline: 24311393]
32. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]
33. Bauermeister JA, Pingel E, Zimmerman M, Couper M, Carballo-Dieguez A, Strecher VJ. Data quality in HIV/AIDS web-based surveys: handling invalid and suspicious data. Field Methods 2012 Apr 26;24(3):272-291. [doi: 10.1177/1525822X12443097]
34. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol 2004 Mar;59(2):93-104. [doi: 10.1037/0003-066X.59.2.93] [Medline: 14992636]
35. Teitcher JE, 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]
36. Konstan JA, Simon Rosser BR, Ross MW, Stanton J, Edwards WM. The story of subject naught: a cautionary but optimistic tale of Internet survey research. J Comput Mediat Commun 2005 Jan;10(2):00. [doi: 10.1111/j.1083-6101.2005.tb00248.x]
37. Grey JA, Konstan J, Iantaffi A, Wilkerson JM, Galos D, Rosser BRS. An updated protocol to detect invalid entries in an online survey of men who have sex with men (MSM): how do valid and invalid submissions compare? AIDS Behav 2015 Oct;19(10):1928-1937. [doi: 10.1007/s10461-015-1033-y] [Medline: 25805443]
38. Mustanski BS. Getting wired: exploiting the Internet for the collection of valid sexuality data. J Sex Res 2001 Nov;38(4):292-301. [doi: 10.1080/00224490109552100]
39. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying multiple submissions in Internet research: preserving data integrity. AIDS Behav 2008 Feb 1;12(6):964-973. [doi: 10.1007/s10461-007-9352-2] [Medline: 18240015]
40. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
41. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
42. Siegel JT, Alvaro EM, Hohman ZP. A dawning recognition of factors for increasing donor registration: The IIFF Model. In: Understanding Organ Donation: Applied Behavioral Science Perspectives. West Sussex, UK: John Wiley & Sons Ltd; 2010:313-330.
43. Kurz RS, Scharff DP, Terry T, Alexander S, Waterman A. Factors influencing organ donation decisions by African Americans: a review of the literature. Med Care Res Rev 2007 Oct;64(5):475-517. [doi: 10.1177/1077558707304644] [Medline: 17881619]

44. Pequegnat W, Rosser BR, Bowen AM, Bull SS, DiClemente RJ, Bockting WO, et al. Conducting Internet-based HIV/STD prevention survey research: considerations in design and evaluation. AIDS Behav 2007 Jul;11(4):505-521. [doi: 10.1007/s10461-006-9172-9] [Medline: 17053853]
45. Steckler A, Linnan L. Process evaluation for public health interventions and research: an overview. In: Process Evaluation for Public Health Interventions and Research. San Francisco, CA: John Wiley & Sons, Inc; 2002:1-23.
46. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]
47. Cummins CO, Prochaska JO, Driskell M, Evers KE, Wright JA, Prochaska JM, et al. Development of review criteria to evaluate health behavior change websites. J Health Psychol 2003 Jan;8(1):55-62. [doi: 10.1177/1359105303008001434] [Medline: 22113900]
48. O'Grady L, Witteman H, Bender JL, Urowitz S, Wiljer D, Jadad AR. Measuring the impact of a moving target: towards a dynamic framework for evaluating collaborative adaptive interactive technologies. J Med Internet Res 2009;11(2):e20 [FREE Full text] [doi: 10.2196/jmir.1058] [Medline: 19632973]
49. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
50. Evers KE. eHealth promotion: the use of the Internet for health promotion. Am J Health Promot 2006;20(4):suppl 1-s7, iii. [Medline: 16555803]
51. Bock B, Graham A, Sciamanna C, Krishnamoorthy J, Whiteley J, Carmona-Barros R, et al. Smoking cessation treatment on the Internet: content, quality, and usability. Nicotine Tob Res 2004 Apr;6(2):207-219. [doi: 10.1080/14622200410001676332] [Medline: 15203794]
52. Danaher BG, Smolkowski K, Seeley JR, Severson HH. Mediators of a successful web-based smokeless tobacco cessation program. Addiction 2008 Oct;103(10):1706-1712 [FREE Full text] [doi: 10.1111/j.1360-0443.2008.02295.x] [Medline: 18715238]
53. Escoffery C, McCormick L, Bateman K. Development and process evaluation of a web-based smoking cessation program for college smokers: innovative tool for education. Patient Educ Couns 2004 May;53(2):217-225. [doi: 10.1016/S0738-3991(03)00163-0]
54. Saunders RP, Evans MH, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. Health Promot Pract 2005 Apr;6(2):134-147. [doi: 10.1177/1524839904273387] [Medline: 15855283]
55. Resnicow K, Baranowski T, Ahluwalia JS, Braithwaite RL. Cultural sensitivity in public health: defined and demystified. Ethn Dis 1999;9(1):10-21. [Medline: 10355471]
56. Callender CO, Miles PV. Obstacles to organ donation in ethnic minorities. Pediatr Transplant 2001 Dec;5(6):383-385. [Medline: 11737760]
57. Callender CO, Miles PV. Minority organ donation: the power of an educated community. J Am Coll Surg 2010 May;210(5):708-715. [doi: 10.1016/j.jamcollsurg.2010.02.037]
58. Arriola K, Robinson DH, Thompson NJ, Perryman JP. Project ACTS: an intervention to increase organ and tissue donation intentions among African Americans. Health Educ Behav 2010 Apr;37(2):264-274 [FREE Full text] [doi: 10.1177/1090198109341725] [Medline: 19858313]
59. Arriola KR, Robinson DH, Perryman JP, Thompson NJ, Russell EF. Project ACTS II: organ donation education for African American adults. Ethn Dis 2013;23(2):230-237 [FREE Full text] [Medline: 23530306]
60. Arriola KR, Powell CL, Thompson NJ, Perryman JP, Basu M. Living donor transplant education for African American patients with end-stage renal disease. Prog Transplant 2014 Dec;24(4):362-370. [doi: 10.7182/pit2014830] [Medline: 25488560]
61. Thorogood M, Coombes Y. Evaluating Health Promotion: Practice and Methods. New York: Oxford University Press; 2010.
62. Shannon DM, Johnson TE, Searcy S, Lott A. Using electronic surveys: advice from survey professionals. Pract Assess Res Eval 2002 Jan;8:1-9.

Abbreviations

ACTS: About Choices in Transplantation and Sharing
CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart
IP: Internet protocol

Edited by G Eysenbach; submitted 19.07.17; peer-reviewed by H Traino, PhD, MPH, K Dwyer; comments to author 17.08.17; revised version received 25.09.17; accepted 29.10.17; published 30.11.17.

Please cite as:
Redmond N, Harker L, Bamps Y, Flemming SSC, Perryman JP, Thompson NJ, Patzer RE, Williams NSD, Arriola KRJ
Implementation of a Web-Based Organ Donation Educational Intervention: Development and Use of a Refined Process Evaluation Model
J Med Internet Res 2017;19(11):e396
URL: http://www.jmir.org/2017/11/e396/
doi:10.2196/jmir.8501
PMID:29191799

Original Paper

Relationship Between the Menstrual Cycle and Timing of Ovulation Revealed by New Protocols: Analysis of Data from a Self-Tracking Health App

Satoshi Sohda[1], MD, PhD; Kenta Suzuki[2,3], PhD; Ichiro Igari[3], PhD

[1]Department of Obstetrics and Gynaecology, Graduate School of Medical Science, University of Tsukuba, Tsukuba, Japan

[2]Biodiversity Conservation Planning Section, Center for Environmental Biology and Ecosystem Studies, National Institute for Environmental Studies, Tsukuba, Japan

[3]Advanced Technology Section, R&D Division, MTI Ltd, Shinjyuku, Japan

Corresponding Author:
Kenta Suzuki, PhD
Biodiversity Conservation Planning Section
Center for Environmental Biology and Ecosystem Studies
National Institute for Environmental Studies
EEF, 3rd Floor
16-2 Onogawa
Tsukuba, 305-8506
Japan
Phone: 81 029 850 2747
Fax: 81 029 850 2747
Email: kenta11514201@gmail.com

Abstract

Background: There are many mobile phone apps aimed at helping women map their ovulation and menstrual cycles and facilitating successful conception (or avoiding pregnancy). These apps usually ask users to input various biological features and have accumulated the menstrual cycle data of a vast number of women.

Objective: The purpose of our study was to clarify how the data obtained from a self-tracking health app for female mobile phone users can be used to improve the accuracy of prediction of the date of next ovulation.

Methods: Using the data of 7043 women who had reliable menstrual and ovulation records out of 8,000,000 users of a mobile phone app of a health care service, we analyzed the relationship between the menstrual cycle length, follicular phase length, and luteal phase length. Then we fitted a linear function to the relationship between the length of the menstrual cycle and timing of ovulation and compared it with the existing calendar-based methods.

Results: The correlation between the length of the menstrual cycle and the length of the follicular phase was stronger than the correlation between the length of the menstrual cycle and the length of the luteal phase, and there was a positive correlation between the lengths of past and future menstrual cycles. A strong positive correlation was also found between the mean length of past cycles and the length of the follicular phase. The correlation between the mean cycle length and the luteal phase length was also statistically significant. In most of the subjects, our method (ie, the calendar-based method based on the optimized function) outperformed the Ogino method of predicting the next ovulation date. Our method also outperformed the ovulation date prediction method that assumes the middle day of a mean menstrual cycle as the date of the next ovulation.

Conclusions: The large number of subjects allowed us to capture the relationships between the lengths of the menstrual cycle, follicular phase, and luteal phase in more detail than previous studies. We then demonstrated how the present calendar methods could be improved by the better grouping of women. This study suggested that even without integrating various biological metrics, the dataset collected by a self-tracking app can be used to develop formulas that predict the ovulation day when the data are aggregated. Because the method that we developed requires data only on the first day of menstruation, it would be the best option for couples during the early stages of their attempt to have a baby or for those who want to avoid the cost associated with other methods. Moreover, the result will be the baseline for more advanced methods that integrate other biological metrics.

(J Med Internet Res 2017;19(11):e391) doi:10.2196/jmir.7468

KEYWORDS

self-tracking; person generated health data; calendar calculation; fertility; menstrual cycle

Introduction

Awareness of one's own fertility is considered important in helping women become pregnant in a shorter period of time [1-3]. Fertility depends on the menstrual (ovarian) cycle, and in each cycle there is a "fertile window" during which women can conceive [1,4,5]. Studies on standard menstrual cycles suggest that the fertile window starts 5 days prior to ovulation and ends on the day of ovulation [4]. This is essentially consistent with the results of statistical studies [1,5] that estimated the day-specific pregnancy rates; the pregnancy rate starts to increase 8 days prior to ovulation, peaks after 6 days (2 days prior to ovulation), and reaches almost 0 at 2-3 days after ovulation. Hence, in order to be aware of the fertility window, it is important for a woman to be able to predict the next ovulation date in the course of her menstrual cycles.

The "calendar method" of predicting the next ovulation date was developed based on the recognition of cycles in the menstrual period and fertility, in which women record their menstrual cycles for family planning [6]. It may have been one of the most widely recorded personal health information details before the spread of mobile phones and computers. On the other hand, the recent popularity of self-tracking tools realized by ubiquitous and wearable technologies has led people to gather various kinds of self-information ranging from financial behaviors to physical activities [7,8]. These technologies are used to "help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge" and are referred to as personal informatics systems [9]. Currently, the classical calendar method of predicting the next ovulation date is integrated into personal informatics systems. There are many mobile phone apps aimed at helping women map their ovulation and menstrual cycles and facilitating successful conception (or avoiding pregnancy) [10,11]. Apps available for these purposes include Ovia Fertility (Ovuline), Glow, OvuView, Ovulation Calendar, Fertility Calendar, My Days, Period Diary, Period Tracker, Maybe Baby, and Fertility Friend [12]. As Lupton [12] mentioned, while these apps are grounded on traditional gynecological knowledge, the advantage that some of them claim is a data analytic approach that can provide greater accuracy than more traditional forms of self-tracking. These apps usually ask users to input various biological features (eg, ovulation, sexual intercourse, basal body temperature, state of cervical mucus, body weight, and the timing of menstrual bleeding).

However, it is not known how to process these features numerically to improve the ovulation prediction error. A mixture of knowledge on biological mechanisms and a statistical approach using the newly enabled biological metrics is promising [13,14], although it is still an open problem. Instead of pursuing the usage of various biological metrics, we believe that more attention should be paid to other aspects of this dataset, that is, its massiveness. It is a remarkable achievement that mobile phone apps have been able to accumulate menstrual cycle data of a vast number of women. Currently, records of more than 10,000 individuals can be a target of statistical analysis. In this study, we start from calendar-based methods that require only the recording of menstruation to predict the ovulation date [15-17]. For many couples, the calendar-based methods are the simplest options of determining the timing of the menstrual cycle [2]. Even within the simple prediction framework, a large amount of data potentially allows us to figure out individual differences better than in traditional understanding [18,19].

This study aims to clarify the above points using data obtained from a commercial women's health care service provided as a mobile phone app in Japan. We extracted approximately 0.1% of all users of the app, resulting in 7043 subjects after data screening. This paper reports on the progressive health data ecosystem in which commercial health care mobile apps generate massive amounts of data. The results of the data analysis give feedback to the app and can be used to improve public health as well.

Methods

Data

We screened 150,000 users who wanted to conceive out of a total of 8,000,000 users of a mobile phone app from a commercial women's health care service, Luna Luna. We used the data of 7043 women who had each recorded at least one menstrual cycle with ovulation date, suggesting that about 5% of women utilize ovulation tests to support their conception. The ovulation date had been determined by one of the methods described in the next paragraph. The total number of cycles was 135,666, and there were 12,731 cycles with an ovulation date. Any cycles in the record that were less than 20 days or greater than 45 days were removed to rule out unnatural cycle length that is due to erroneous or defective input. About 57% (7285) of the cycles had more than 8 records of past menstrual cycles after the screening, which allowed us to analyze the relationship between past and future menstrual cycles. The age distribution of the 7043 women ranged from 20-45 years with a mean of 32.94 years (95% CI 32.04-33.85), which is slightly higher than the mean maternal age of Japanese women at the time of first birth (which was 30.1 years in 2010).

Each user's personal records consisted of the dates when they recognized menstrual bleeding (onset of menstruation) and the dates when they detected ovulation. In the log file, an identifier is attached to each ovulation record to distinguish its basis (clinical diagnosis/ovulation test kit/other reliable method). In our analysis, only clinical diagnosis based (31%) and ovulation test kit based (54%) ovulation records were used. Luna Luna does not ask women to record which clinical diagnostic test they used to determine the ovulation date. However, it is noted that the ovulation day in Japan is commonly determined by ultrasound scanning and occasionally with testing of blood luteinizing hormone or estrogen level.

The Luna Luna data are the property of MTI Ltd (Shinjyuku, Tokyo, Japan). The authors (KS and II) are employees of MTI Ltd and are permitted to access the Luna Luna data server. The information security committee of MTI Ltd concluded that this study does not require approval by an ethics committee because the data are anonymized appropriately; the data server used in the study is a backup of the original data server, on which anonymous IDs are placed on personally identifiable information. Hence, it was impossible for the authors to access personally identifiable information, which was controlled separately. Consent for data use and information acquisition was obtained from Luna Luna users, as stipulated in the terms of use.

Luna Luna

Luna Luna is a total health care service for female mobile phone users in Japan. Luna Luna offers its users predictions of menstrual cycles, fertility, ovulation, and related health care information, based on user-inputted personal records that are sent and stored in its data server. The data are securely stored separately from personally identifiable information. Luna Luna has been provided as a commercial service for more than 10 years since the year 2000. Luna Luna has 7 million subscribers as of 2016 and occupies a leading position among mobile health care services for female users in the Japanese market.

Menstrual Cycles, Timing of Ovulation, Follicular Phases, and Luteal Phases

We express the records of the first day of menstruation of woman i as,

$$M_i=(m_{i1}, m_{i2},...,m_{iT}),$$

where m_{i1} is the first day of the most recent menstruation of woman i $(i=1, 2,...,N)$, m_{i2} is the first day of her second most recent menstruation, and so on. Then, we defined Ci as a series of menstrual cycle lengths of woman i by,

$$C_i=(c_{i1},...,c_{iT-1})=(D(m_{i1}, m_{i2}),...,D(m_{iT-1}, m_{iT})),$$

where c_{ij} denotes the j th most recent menstrual cycle of woman i and $D(m_{ij}, m_{ij+1})$ denotes a function that gives the number of elapsed days between m_{ij+1} and m_{ij}. For simplicity of notation, we define the mean length of the menstrual cycles of woman i over the j th to j' th cycles as $c_i*(j, j')=\Sigma_{t=j,...,j'}c_{it}/(j'-j+1)$.

F_i is the series of follicular phase lengths of woman i, where each follicular phase length, f_{ij}, is defined as follows: assuming we have records of ovulation $o_{ij}(j=1, 2, ...)$ between m_{ij} and m_{ij+1},

$$F_i=(..., f_{ij}, ...)=(..., D(o_{ij}, m_{ij+1}), ...).$$

Thus, the timing of ovulation is f_{ij} days after the day of the previous menstruation m_{ij+1}. Similarly, the series of luteal phase lengths of woman i, L_i, where each luteal phase length is l_{ij}, is defined as,

$$L_i=(..., l_{ij}, ...)=(..., D(m_{ij}-1, o_{ij}+1), \ldots).$$

Here, $o_{ij}+1$ is the day after ovulation and $m_{ij}-1$ is the day before the next menstruation. The timing of ovulation is $c_{ij}-l_{ij}-1$ days after the first day of the previous menstruation m_{ij+1}. It should be mentioned that the length of records varied among the women. We used T_i to indicate the length of records of woman i.

We investigated the relationships between the length of menstrual cycles and the length of the follicular phases or that of the luteal phases. We also analyzed the relationships between the mean length of past menstrual cycles and the length of the follicular phases or that of the luteal phases because prediction of ovulation date requires an unknown length of the next menstrual cycle.

Calendar Calculations

We evaluated the relevance of three calendar-based methods using our data. The first was the Ogino method [15], which assumes a fixed length of the luteal phase of 14 days and predicts the ovulation date as $[c^*]-15$ days after the onset of the previous menstruation for a particular woman when the mean length of her menstrual cycles is c^* (here $[c^*]$ represents c^* rounded down to the nearest integer). This implies that the length of the follicular phase in the next cycle is $[c^*]-15$. The Ogino method is the most widespread calendar-based method relied on by Japanese women. The second method is the method proposed by Lamprecht and Grummer-Strawn [16], which assumes that the length of the next follicular phase is $[c^*/2]$, and thus this method predicts a woman's next ovulation date as $[c^*/2]$ days after the onset of the previous menstruation. Here, we call this method as the half cycle length (HCL) method. Because of its simplicity, we chose the HCL method over other calendar-based methods that reflect individual differences in the length of the luteal phase in a menstrual cycle. The third method is the method that was developed in this study and predicts a woman's next ovulation date as $f_i(c^*)$ days after the onset of the previous menstruation. As explained in the next section, f_i is a linear function that is optimized by using the relationship between the follicular phase length and c^*. Hence, we call the third method as the Optimized method.

To analyze the relationship between the timing of ovulation and the mean length of past menstrual cycles, we evaluated the three prediction models for follicular phase length, that is, μ_{Ogino}, μ_{HCL}, and μ_{OPT}, which predict the length of the next follicular phase as $\mu_{Ogino}(C_i, j, k)=[c_{ij}*(k)]-15$, $\mu_{HCL}(C_i, j, k)=[c_{ij}*(k)/2]$, and $\mu_{OPT}(C_i, j, k)=[f_{ij}(c_{ij}*(k))]$, respectively. Here, $c_{ij}*(k)$ is defined as $c_{ij}*(k)=c_i*(j+1. J+k)$. It should be noted that if $k=1$, $c_{ij}*(k)$ is identical to the nearest cycle length, c_{ij+1}.

To predict the timing of ovulation, we used the results of least square fitting between mean cycle length and follicular phase length. The prediction performance of the timing of ovulation using the obtained model, μ_{OPT}, and the prediction performances using the Ogino and HCL methods, μ_{Ogino} and μ_{HCL}, respectively, were compared with different allowable prediction error levels, $|\mu-f_{ij}| \leq 0, 1, 2$.

Linear Models

In our analysis, we used a linear model to describe the relationship between an explanatory variable x and a response

variable *y*. For data that consist of multiple data points from each individual, linear models are generally categorized into two types: fixed effect models and random effect models [20]. A fixed effect model is formalized as follows:

$$y_{ij}=\alpha_1\delta_{i1}+\ldots+\alpha_n\delta_{in}+x_{ij}\beta+\varepsilon_{ij},$$

$$\delta_{ij}=1 \text{ if } i=j \text{ else } 0,$$

where δ_{ij} s are dummy variables. The least square estimate of parameters including dummy variables is obtained as,

$$\beta_{lsdv}=\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,Ti}(x_{ij}-x_i^*)(x_{ij}-x_i^*)/\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,Ti}(x_{ij}-x_i^*)(y_{ij}-y_i^*),$$

$$\alpha_i=y_i^*-x_i^*\beta_{lsdv},$$

where $x_i^*=1/T_i\sum_{t=1,\ldots,Ti}x_{it}$ and $y_i^*=1/T_i\sum_{t=1,\ldots,Ti}y_{it}$. Hence, in this model, different women have different α_i for the relationship between *x* and *y*, and there is a correlation between α and x. On the other hand, in random effect models, the random component of α_i s is included in the random variable μ_{ij} as follows:

$$y_{ij}=\alpha+x_{ij}\beta+\mu_{ij}=x_{ij}'\beta'+\mu_{ij},$$

Here, we assumed that $x_{ij}'=(1, x_{ij})$ and $\beta'=(\alpha, \beta)$. A pooled ordinary least square estimate, $\beta_p'=(\alpha_p, \beta_p)$, is obtained as,

$$\beta_p'=\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,Ti}x_{it}'y_{it}/\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,Ti}x_{it}'^2.$$

However, the result underestimates the covariance structure in μ_{ij} s, which is described as a matrix,

$$\Omega=\{\omega_{ij}\}_{i,\,j=1,\ldots,N},$$

$$\omega_{ij}=\xi_\alpha^2+\xi^2 \text{ if } i=j \text{ else } \xi_\alpha^2.$$

Using the result of pooled ordinary least square, the generalized least square estimate of parameters that include the effect of covariance, $\beta_{gls}'=(\alpha_{gls}, \beta_{gls})$, is obtained as,

$$\beta_{gls}'=\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,Ti}\sum_{s=1,\ldots,Ti}\omega_{ts}^{(-1)}x_{it}'y_{it}/\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,Ti}\sum_{s=1,\ldots,Ti}\omega_{ts}^{(-1)}x_{it}'^2,$$

where $\omega_{ts}^{(-1)}$ is an element of $\Omega^{(-1)}$, which is the inverse matrix of Ω. To obtain ξ_α and ξ , we first calculate,

$$\xi_\mu=1/(NT-p)\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,Ti}\mu_{it},$$

and

$$\xi_\alpha=1/\{NT(T-1)/2-p\}\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,Ti-1}\sum_{s=t+1,\ldots,Ti}\mu_{it}\mu_{is},$$

then,

$$\xi=\xi_\mu-\xi_\alpha$$

In this paper, we used the Hausman test to determine which of the models better explains the data. In the Hausman test, the percentile value of $H=(\beta_{gls}-\beta_{lsdv})^2$ in a X^2 distribution with one degree of freedom is calculated, and a random effect model is rejected if the *P* value is greater than .05.

Results

In our dataset, the mean (95% confidence interval) of the menstrual cycle length, the follicular phase length, and the luteal phase length of the 7043 women over all cycles was 29.76 (24-38), 14.84 (10-23), and 13.91 (10-19) days, respectively. Both the length of the follicular phases and of the luteal phases had a positive correlation with the length of the menstrual cycles (Table 1). The Pearson correlation coefficient between the length of the menstrual cycles and the length of the follicular phases or of the luteal phases was .75 ($P<.001$) and .37 ($P<.001$), respectively. Hence, both the follicular phase length and luteal phase length had significant positive correlations with the menstrual cycle length. For analysis of the relationship between the length of the menstrual cycles and the length of the follicular phases or of the luteal phases, we applied the random effect model because the *P* value of the Hausman test was .001. The coefficient (α, β) of the generalized least square estimate was (.501, -.088) for the follicular phase length and (.466, .088) for the luteal phase length.

We then investigated the relationship between the mean length of past cycles and the cycle length, follicular phase length, and luteal phase length of the next menstrual cycle. Table 2 shows the number of cycles having enough records to calculate $c_{ij}*(k)$ for each number of k. For example, there were 11,640 cycles with at least one previous cycle, and there were 7285 cycles having records of 8 past cycles. Table 3 shows the Pearson correlation coefficients between $c_{ij}*(k)$ and the next menstrual cycle length, follicular phase length, and luteal phase length. Both the next cycle length and the follicular phase length had strong correlations with the mean cycle length. Only a weak correlation was found between the mean cycle length and the luteal phase length, although it was statistically significant. We applied the random effect model because the *P* value of the Hausman test was <.050 for all cases (Table 4). The coefficient (α, β) of the generalized least square estimate was similar to that calculated for the actual cycle length. In summary, the menstrual cycle length had positive correlations with both the follicular phase length and luteal phase length, although the correlation was less strong with luteal phase length.

The prediction performance of the timing of ovulation using the obtained model, μ_{OPT}, was compared with that of μ_{Ogino} and μ_{HCL}. μ_{OPT} outperformed μ_{Ogino} when a woman's mean cycle length was shorter than 27 days or longer than 31 days (Figure 1). μ_{OPT} outperformed μ_{HCL} when a woman's mean cycle length was less than 28 days. As for the mean accuracy over different mean cycle lengths (Figure 2), μ_{OPT} outperformed μ_{Ogino} in all cases. The prediction performances of μ_{OPT} and μ_{HCL} were similar when only small numbers of cycles were available to calculate the mean cycle length, whereas μ_{OPT} showed an advantage with increasing values of *k*.

Table 1. Relationship between cycle length, follicular phase length, and luteal phase length.

Cycle length, days	Cycles, n	Mean follicular phase length, days	95% CI	Mean luteal phase length, days	95% CI
23	120	10.5	(7-15)	11.5	(7-15)
24	324	11.1	(7-15)	11.9	(8-16)
25	657	11.5	(8-16)	12.5	(8-16)
26	1065	12.1	(9-16)	12.9	(9-16)
27	1407	12.7	(10-16)	13.3	(10-16)
28	1637	13.4	(10-17)	13.6	(10-17)
29	1516	14.1	(11-17)	13.9	(11-17)
30	1392	15	(11-19)	14	(10-18)
31	1144	15.7	(12-20)	14.3	(10-18)
32	875	16.4	(12-20)	14.6	(11-19)
33	707	17.4	(13-21)	14.6	(11-19)
34	557	18.1	(12-23)	14.9	(10-21)
35	395	18.8	(11-26)	15.2	(8-23)
36	300	19.6	(12-25)	15.4	(10-23)
37	204	20.1	(12-25)	15.9	(11-24)
38	162	21.5	(14-28)	15.5	(9-23)
39	145	21.6	(12-29)	16.4	(9-26)
40	124	22.3	(12-29)	16.7	(10-27)

Table 2. Number of cycles having records of k past cycles.

Mean cycle length, days	k=1	2	3	4	5	6	7	8
24	176	168	117	121	90	91	67	68
25	438	496	384	385	325	313	282	272
26	878	918	805	798	694	700	625	608
27	1253	1291	1194	1222	1129	1083	1003	911
28	1569	1529	1507	1440	1382	1293	1196	1154
29	1620	1541	1472	1361	1306	1270	1178	1078
30	1396	1328	1292	1238	1181	1074	998	934
31	1101	1024	1036	979	964	895	860	785
32	862	856	867	728	715	647	624	560
33	721	637	564	569	520	463	429	378
34	532	448	461	371	362	301	276	225
35	377	353	314	253	226	188	195	174
36	290	214	194	179	141	110	88	76
37	219	160	125	90	74	65	47	42
Total	11,640	11093	10421	9804	9154	8529	7898	7285

Table 3. Pearson correlation coefficients between mean cycle length and next cycle length, follicular phase length, or luteal phase length (P<.001 for all cases).

	k=1	2	3	4	5	6	7	8
Next cycle length, days	.543	.574	.584	.589	.589	.593	.592	.592
Follicular phase length, days	.506	.526	.537	.547	.547	.546	.548	.547
Luteal phase length, days	.109	.129	.138	.135	.137	.144	.140	.140

Table 4. Results of least square fitting and the Hausman test.

	k=1	2	3	4	5	6	7	8
Follicular phase length								
A	.528	.528	.527	.526	.526	.525	.525	.523
B	.039	.017	.012	.003	–.002	–.010	–.011	–.011
P value	.001	.001	.001	.001	.001	.001	.001	.001
Luteal phase length								
A	.465	.467	.470	.471	.471	.472	.472	.473
B	.039	.017	.012	.003	–.002	–.010	–.011	–.011
P value	.001	.001	.001	.001	.001	.001	.001	.001

Figure 1. Accuracy of the Ogino, HCL, and the Optimized methods in predicting the day of the next ovulation date with allowable prediction error=0 (a-d), 1 (e-h), and 2 (i-j) for different mean cycle lengths (solid line, dashed line, and dotted line indicate the Ogino, HCL, and the Optimized methods, respectively).

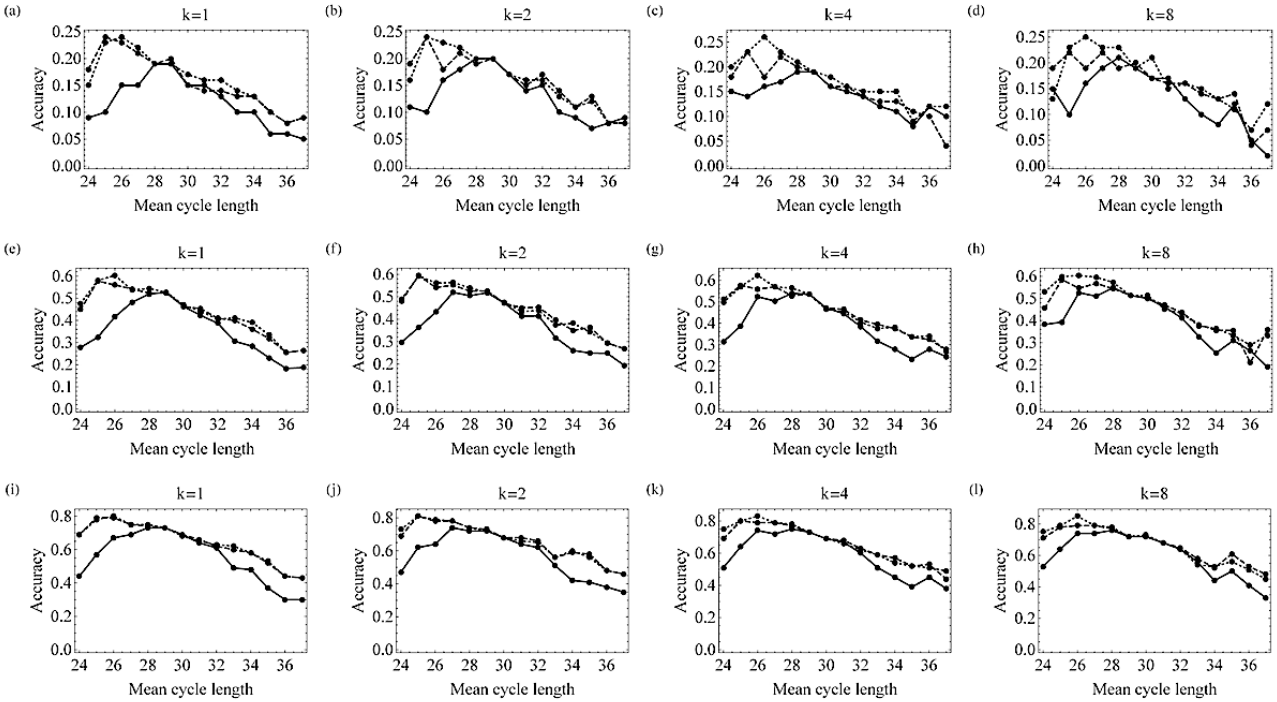
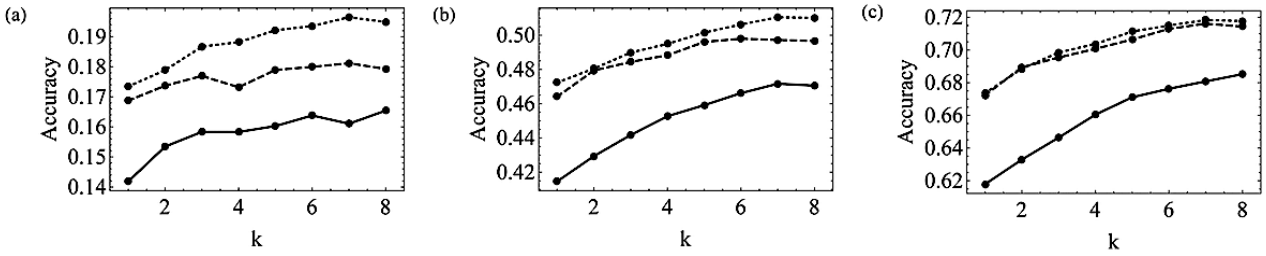


Figure 2. Mean accuracy over different mean cycle lengths with allowable prediction error=0 (a), 1 (b), and 2 (c) (solid line, dashed line, and dotted line indicate the Ogino, HCL, and the Optimized methods, respectively).



Discussion

Principal Findings

In this paper, using the data obtained from the Luna Luna service, we evaluated how menstrual cycle length is related to luteal phase length and follicular phase length. As suggested in a previous study [21], the correlation between length of menstrual cycles and length of follicular phases was stronger than the correlation between length of menstrual cycles and length of luteal phases (Table 1). Furthermore, there was a positive correlation between the lengths of past and future menstrual cycles (Tables 3 and 4); this suggests that there is regularity in menstrual cycles as claimed by Creinin et al [22]. A strong positive correlation was also found between the mean length of past cycles and the mean length of follicular phases. Thus, the follicular phase length can be mostly estimated from each woman's menstrual cycle length. However, the correlation between mean cycle length and luteal phase length was also statistically significant. Our results showed that the random effect model was relevant for modeling the relationship between menstrual cycle length and follicular phase length as well as the relationship between menstrual cycle length and luteal phase length. Furthermore, the random effect model was also relevant for modeling the relationships between mean length of past menstrual cycles and mean length of follicular phases or luteal phases. These results suggest that these relationships are shared among all women without being influenced by personal factors. The existing calendar-based methods (Ogino and HCL methods) did not explain our data well. Against the assumption of the Ogino method, the weak positive correlation between mean cycle length and luteal phase length (Tables 3 and 4) suggests that the luteal phase does not have a constant length among women with mean menstrual cycles of different lengths. This is why the Ogino method was less accurate when a woman's mean cycle length was shorter than 27 days or longer than 32 days (Figure 1). The HCL method was less accurate when a woman's mean cycle length was less than 28 days (Figure 1). By taking the average over different mean cycle lengths, the accuracy of the Ogino method was worse than that of the other methods as well (Figure 2). The accuracy of the HCL method was close to that of the Optimized method when there were only a small number of cycles available to calculate the mean cycle length (Figure 2). The Optimized method showed better performance with increasing numbers of available cycles. The Optimized method outperformed the HCL method in explaining the relationship between the mean cycle length and the follicular phase length when a large number of past cycles were available to calculate the mean cycle length. These results showed that the performance of the Optimized method was equivalent to or better than that of the HCL method. Hence, we recommend using the Optimized method to predict the timing of ovulation from the mean length of menstrual cycles if these data are available.

The novelty of our findings is essentially accounted for by the large number of participants. The Ogino method was developed based on the anatomical observation of ovarian follicles of 81 women with cycle lengths of between 23 and 45 days [15]. They reported that there was a variable luteal phase length (12-16 days), while the number of subjects studied was not large enough to identify the more detailed relationship between the lengths of the menstrual cycle, follicular phase, and luteal phase. Three studies with more data reported the mean (95% CI) of the follicular phase length and the luteal phase length as 16.5 (9-23) and 12.4 (8-17) [21], 15.5 (9-22) and 12.6 (9-17) [23], and 15.0 (2-21) and 13.5 (7-20) days [24], respectively. These studies reported only the mean value over all women, except for Fehring et al [21] who reported positive correlations between menstrual cycle length and follicular phase length or luteal phase length. However, their analysis was based on only 1060 cycles in 141 women. Lamprecht and Grummer-Strawn [16] suggested that follicular phase length was better estimated by *c* */2 based on an analysis of 7514 cycles in 1062 women. In contrast, our study analyzed more than 65,000 cycles, which is 61 and 8.5 times greater than what was used in the studies of Fehring et al [21] and Lamprecht and Grummer-Strawn [16], respectively. This study allowed us to capture the relationships between length of menstrual cycle, follicular phase, and luteal phase in more detail.

Recently, mobile health information technology—known as "digital health," "eHealth," or "mHealth"—has been used in medicine and public health in various ways [25-28]. With the recent spread of smartphones and other mobile devices (eg, 77% of US adults [29], 62.6% of Japanese [30], and 62% of Chinese [31] own smartphones), digital health technologies have begun to be adapted for a wide variety of purposes. The mHealth information technologies are one of the origins of the advanced movement called "Quantified Self", which stresses the role of patients or consumers in medicine and public health [18,19,32]. Quantified-self refers to an individual who is engaged in the self-tracking of any kind of biological, physical, behavioral, or environmental information [19]. These movements are now spreading among people who were not familiar with such technologies before [7]. For example, 60% of US adults are currently tracking their weight, diet, or exercise routine, and 33% are monitoring other factors such as blood sugar, blood

pressure, headaches, or sleep patterns [33]. There are 165,000 mobile phone health apps available on the market [34]. The self-tracked health data are regarded as the key to realizing personalized medicine and health maintenance [19].

On the other hand, there are several concerns about these technologies. First, Lupton [35] argued that they might allow our personal health data be used as "both an object of surveillance and persuasion" [35]. The author in particular claimed that the sharing of health data on social networking services (SNSs) would result in "public surveillance" where everyone would be a subject of surveillance by others. However, it is less beneficial for women to share their menstrual cycle records on SNSs because menstrual cycles strongly depend on biological and physiological factors. Furthermore, issues about menstrual cycles are "sensitive" personal information that would not likely be a topic of conversation among friends. Hence, it is unlikely that women would start sharing information on their menstrual cycles on SNSs. On the other hand, health informatics systems can integrate these data both systematically and anonymously and provide feedback knowledge at a scale that is not achieved by any person-to-person communications. These systems allow women to maximize the benefit of sharing data on their menstrual cycles (or other sensitive health information) without publicizing the data themselves. Sharon [36] argued that self-tracking health technologies reduce phenomena to numbers and "that this simultaneously displaces other, non-quantifiable yet highly insightful means of knowing and expression". For example, a small percentage of women are aware that they have somatic symptoms around the time of ovulation, for example, ovulation pain. This self-awareness would help women recognize the timing of ovulation. However, not all women have or are aware of these symptoms. Hence, there is a benefit of summarizing regularity behind menstrual cycles in a simple rule and sharing it as social knowledge [17]. We claim that mHealth technology puts this knowledge sharing onto a new stage because it can relax the restriction that the rules must be simple enough to be handled by anyone. Of course, there is loss of information in the rule extraction process. The providers of health informatics systems should take this point seriously and should aim at designing their systems [7-9] so that users can maintain an appropriate distance between their body and its data representation [35,37]. For this purpose, the interaction between a system and its users should be studied in terms of user behavior to find a better way of presenting predictive performance, and these insights should be incorporated into the design of the system [37-39]. Most people are still sensitive about unintended use of data by for-profit companies. In a recent survey on attitudes toward personal health care data [40], more than 80% of the respondents answered that they were willing to share health and medical information with a personal physician or health care provider (88%) and with nonprofit research organizations (84%). Only 24% answered that they were willing to share their data with for-profit companies. Nevertheless, it is worth noting that 88% answered that their motivation for sharing health information was to make new health discoveries. With increasing numbers of people contributing their health data to analyzable datasets, health information data will not only empower companies and consumers but also benefit health and social systems [18].

Limitations

Our analysis lacked complete profile data for all subjects and the dataset had inevitable selection bias. Moreover, menstrual bleeding started in the middle of the night in some cases—the date taken as the menstrual start day depended on the user's subjective choice. However, we believe that having a large-scale dataset available to carry out investigations on women's health overcomes such limitations.

Conclusions

Our study demonstrated how the present calendar methods of predicting the ovulation date were improved by the better grouping of women, which can be supported statistically only with massive numbers of subjects. Because calendar methods require only the timing of menstruation, one of the calendar methods would be the best option for couples during the early stages of their attempt to have a baby or for those who want to avoid the cost associated with other methods. Moreover, the results of the calendar method will be the baseline for more advanced methods that integrate other biological metrics. The mobile phone based health care services are very efficient in obtaining large datasets because they offer easy ways for users to input and manage their personal data. This recently enabled data collection framework is complementary to existing well-controlled experimental methods and will contribute to the testing of medical hypotheses that previously could not be studied due to insufficient numbers of subjects. Users benefit from newly developed medical knowledge by using mobile phone based services without the need to learn intricate calculations. For medical personnel and researchers, the records accumulated by these commercial services can be a useful source of data for analysis after appropriate anonymity processing. Thus, in the mobile phone based services that aim at facilitating conception (or contraception), medical specialists and users form a knowledge-improving cycle that can provide quick feedback to the users from the emerging analysis results. Such systems, including other mobile phone health care services, are strongly expected to contribute to comprehensive health care for people of all ages.

Acknowledgments

We would like to thank Cornelis B Lambalk for valuable comments and suggestions. MTI Ltd provided support in the form of salaries for authors KS and II but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

The study idea was conceived by all authors. The study protocol and analysis were developed by KS and II. The first draft was written by KS and II. SS participated in the revision of the paper and critical discussions. All authors approved the final draft.

Conflicts of Interest

KS and II are employees of MTI Ltd. In addition, KS and II have a patent (Japan Patent Office, No. P5998307) pending.

References

1. Dunson DB, Colombo B, Baird DD. Changes with age in the level and duration of fertility in the menstrual cycle. Hum Reprod 2002 May;17(5):1399-1403. [Medline: 11980771]
2. Stanford JB, White GL, Hatasaka H. Timing intercourse to achieve pregnancy: current evidence. Obstet Gynecol 2002 Dec;100(6):1333-1341. [Medline: 12468181]
3. Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. N Engl J Med 1995 Dec 07;333(23):1517-1521. [doi: 10.1056/NEJM199512073332301] [Medline: 7477165]
4. Mihm M, Gangooly S, Muttukrishna S. The normal menstrual cycle in women. Anim Reprod Sci 2011 Apr;124(3-4):229-236. [doi: 10.1016/j.anireprosci.2010.08.030] [Medline: 20869180]
5. Colombo B, Masarotto G. Daily Fecundability. DemRes 2000 Sep 06;3:5 [FREE Full text] [doi: 10.4054/DemRes.2000.3.5]
6. Spieler JM, Collins WP. Potential fertility - defining the window of opportunity. J Intern Med Res 2001;29(1_suppl):3A-13A [FREE Full text] [doi: 10.1177/14732300010290S102]
7. Rapp A, Cena F. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. International Journal of Human-Computer Studies 2016 Oct;94:1-17 [FREE Full text] [doi: 10.1016/j.ijhcs.2016.05.006]
8. Epstein DA, Ping A, Fogarty J, Munson SA. A lived informatics model of personal informatics. 2015 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2015; Osaka, Japan p. 731-742.
9. Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. 2010 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2010; Atlanta, GA p. 557-566 URL: https://pdfs.semanticscholar.org/f90d/ffcab9850a03a4733abb31c123f244ff5f6c.pdf
10. Simmons RG, Shattuck DC, Jennings VH. Assessing the Efficacy of an App-Based Method of Family Planning: The Dot Study Protocol. JMIR Res Protoc 2017 Jan 18;6(1):e5. [doi: 10.2196/resprot.6886]
11. Mangone ER, Lebrun V, Muessig KE. Mobile Phone Apps for the Prevention of Unintended Pregnancy: A Systematic Review and Content Analysis. JMIR Mhealth Uhealth 2016 Jan 19;4(1):e6 [FREE Full text] [doi: 10.2196/mhealth.4846] [Medline: 26787311]
12. Lupton D. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. Cult Health Sex 2015;17(4):440-453. [doi: 10.1080/13691058.2014.920528] [Medline: 24917459]
13. Pallone SR, Bergus GR. Fertility awareness-based methods: another option for family planning. J Am Board Fam Med 2009;22(2):147-157 [FREE Full text] [doi: 10.3122/jabfm.2009.02.080038] [Medline: 19264938]
14. Fehring RJ. New low- and high-tech calendar methods of family planning. J Midwifery Womens Health 2005;50(1):31-38. [doi: 10.1016/j.jmwh.2004.07.001] [Medline: 15637512]
15. Ogino K. Uber den konzeptionstermin des weibes und seine anwendung in der praxis. Zentralbl Gynakol 1932;56:721-732.
16. Lamprecht VM, Grummer-Strawn L. Development of new formulas to identify the fertile time of the menstrual cycle. Contraception 1996 Dec;54(6):339-343. [Medline: 8968662]
17. Arévalo M, Jennings V, Sinai I. Efficacy of a new method of family planning: the Standard Days Method. Contraception 2002 May;65(5):333-338. [Medline: 12057784]
18. Swan M. Health 2050: The realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen. Journal of Personalized Medicine 2012;2:93-118. [doi: 10.3390/jpm2030093]
19. Swan M. The quantified self: Fundamental disruption in big data science and biological discovery. Big Data 2013;1:85-99. [doi: 10.1089/big.2012.0002] [Medline: 27442063]
20. Schmidheiny K, Basel U. Panel data: fixed and random effects. In: Short Guides to Microeconometrics. Basel, Switzerland: Unversitat Basel; 2011:2-7.
21. Fehring R, Schneider M, Raviele K. Variability in the phases of the menstrual cycle. J Obstet Gynecol Neonatal Nurs 2006;35(3):376-384. [doi: 10.1111/j.1552-6909.2006.00051.x] [Medline: 16700687]
22. Creinin MD, Keverline S, Meyn LA. How regular is regular? An analysis of menstrual cycle regularity. Contraception 2004 Oct;70(4):289-292. [doi: 10.1016/j.contraception.2004.04.012] [Medline: 15451332]
23. France JT, Graham FM, Gosling L, Hair P, Knox BS. Characteristics of natural conceptual cycles occurring in a prospective study of sex preselection: fertility awareness symptoms, hormone levels, sperm survival, and pregnancy outcome. International Journal of Fertility 1991;37:244-255.

24. World Health Organization. A prospective multicentre study of the ovulation method of natural family planning. IV. The outcome of pregnancy. World Health Organization. Fertil Steril 1984 Apr;41(4):593-598. [Medline: 6705935]
25. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
26. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
27. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
28. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
29. Pew Research Center. Device ownership over time. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2017-02-03] [WebCite Cache ID 6nzTpWmFv]
30. Economic Research Office, General Policy Division, Information and Communications Policy Bureau. White Paper. Information and Communications in Japan. Tokyo: Ministry of Internal Affairs and Communications, Japan; 2015.
31. Hsu J, Liu D, Yu Y, Zhao HT, Chen ZR, Li J, et al. The Top Chinese Mobile Health Apps: A Systematic Investigation. J Med Internet Res 2016 Aug 29;18(8):e222 [FREE Full text] [doi: 10.2196/jmir.5955] [Medline: 27573724]
32. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Dec;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
33. Fox S, Duggan M. Mobile health 2012. Washington, DC: Pew Internet & American Life Project; 2012.
34. IMS Institute for Healthcare Informatics. Availability and profile of consumer mHealth apps URL: https://www.imshealth.com/files/web/IMSH%20Institute/Reports/Patient%20Adoption%20of%20mHealth/IIHI-Patient-Adoption-mhealth-Exhibits-Full.pdf [accessed 2017-02-03] [WebCite Cache ID 6nzUmYHJS]
35. Lupton D. M-health and health promotion: The digital cyborg and surveillance society. Soc Theory Health 2012 Jun 27;10(3):229-244 [FREE Full text] [doi: 10.1057/sth.2012.6]
36. Sharon T. Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. Philos Technol 2016 Apr 18;30(1):93-121. [doi: 10.1007/s13347-016-0215-5]
37. Schüll ND. Data for life: Wearable technology and the design of self-care. BioSocieties 2016 Oct 13;11(3):317-333. [doi: 10.1057/biosoc.2015.47]
38. Bentley F, Tollmar K, Stephenson P, Levy L, Jones B, Robertson S, et al. Health Mashups. ACM Trans Comput-Hum Interact 2013 Nov 01;20(5):1-27. [doi: 10.1145/2503823]
39. Rapp A, Tirassa M. Know Thyself: A Theory of the Self for Personal Informatics. Human–Computer Interaction 2017 Jan 27;32(5-6):335-380. [doi: 10.1080/07370024.2017.1285704]
40. Pickard K, Swan M. Big desire to share big health data: A shift in consumer attitudes toward personal health information. 2014 Presented at: AAAI Spring Symposium; 2014; Palo Alto, CA p. 51-58 URL: https://www.aaai.org/ocs/index.php/SSS/SSS14/paper/viewFile/7765/7783

Abbreviations

HCL: half cycle length
OPT: optimal
SNS: social networking service

Edited by G Eysenbach; submitted 07.02.17; peer-reviewed by K Blake, A Rapp; comments to author 20.04.17; revised version received 11.08.17; accepted 08.09.17; published 27.11.17.

Please cite as:
Sohda S, Suzuki K, Igari I
Relationship Between the Menstrual Cycle and Timing of Ovulation Revealed by New Protocols: Analysis of Data from a Self-Tracking Health App
J Med Internet Res 2017;19(11):e391
URL: http://www.jmir.org/2017/11/e391/
doi:10.2196/jmir.7468
PMID:29180346

Original Paper

Implementation Intention and Reminder Effects on Behavior Change in a Mobile Health System: A Predictive Cognitive Model

Peter Pirolli[1], PhD; Shiwali Mohan[2], PhD; Anusha Venkatakrishnan[2], PhD; Les Nelson[2], MSc; Michael Silva[2], BSc; Aaron Springer[3], BSc (Hons)

[1]Institute for Human and Machine Cognition, Pensacola, FL, United States

[2]Palo Alto Research Center, Palo Alto, CA, United States

[3]University of California, Santa Cruz, Santa Cruz, CA, United States

Corresponding Author:
Peter Pirolli, PhD
Institute for Human and Machine Cognition
40 S. Alcanaz St
Pensacola, FL, 32502
United States
Phone: 1 850 202 4462
Email: ppirolli@ihmc.us

Abstract

Background: Implementation intentions are mental representations of simple plans to translate goal intentions into behavior under specific conditions. Studies show implementation intentions can produce moderate to large improvements in behavioral goal achievement. Human associative memory mechanisms have been implicated in the processes by which implementation intentions produce effects. On the basis of the adaptive control of thought-rational (ACT-R) theory of cognition, we hypothesized that the strength of implementation intention effect could be manipulated in predictable ways using reminders delivered by a mobile health (mHealth) app.

Objective: The aim of this experiment was to manipulate the effects of implementation intentions on daily behavioral goal success in ways predicted by the ACT-R theory concerning mHealth reminder scheduling.

Methods: An incomplete factorial design was used in this mHealth study. All participants were asked to choose a healthy behavior goal associated with eat slowly, walking, or eating more vegetables and were asked to set implementation intentions. N=64 adult participants were in the study for 28 days. Participants were stratified by self-efficacy and assigned to one of two reminder conditions: reminders-presented versus reminders-absent. Self-efficacy and reminder conditions were crossed. Nested within the reminders-presented condition was a crossing of frequency of reminders sent (high, low) by distribution of reminders sent (distributed, massed). Participants in the low frequency condition got 7 reminders over 28 days; those in the high frequency condition were sent 14. Participants in the distributed conditions were sent reminders at uniform intervals. Participants in the massed distribution conditions were sent reminders in clusters.

Results: There was a significant overall effect of reminders on achieving a daily behavioral goal (coefficient=2.018, standard error [SE]=0.572, odds ratio [OR]=7.52, 95% CI 0.9037-3.2594, P<.001). As predicted by ACT-R, using default theoretical parameters, there was an interaction of reminder frequency by distribution on daily goal success (coefficient=0.7994, SE=0.2215, OR=2.2242, 95% CI 0.3656-1.2341, P<.001). The total number of times a reminder was acknowledged as received by a participant had a marginal effect on daily goal success (coefficient=0.0694, SE=0.0410, OR=1.0717, 95% CI −0.01116 to 0.1505, P=.09), and the time since acknowledging receipt of a reminder was highly significant (coefficient=−0.0490, SE=0.0104, OR=0.9522, 95% CI −0.0700 to −0.2852], P<.001). A dual system ACT-R mathematical model was fit to individuals' daily goal successes and reminder acknowledgments: a goal-striving system dependent on declarative memory plus a habit-forming system that acquires automatic procedures for performance of behavioral goals.

Conclusions: Computational cognitive theory such as ACT-R can be used to make precise quantitative predictions concerning daily health behavior goal success in response to implementation intentions and the dosing schedules of reminders.

(J Med Internet Res 2017;19(11):e397) doi:10.2196/jmir.8217

KEYWORDS

mobile applications; models, theoretical; habits

Introduction

Background

Mobile health (mHealth) systems provide new opportunities to provide precise individualized just-in-time interventions to support behavior change [1]. mHealth provides a path for translating evidence-based interventions (EBIs) onto delivery systems that are replicable, scalable, and sustainable, with great economies of scale for health care delivery [2]. mHealth also provides new opportunities to study and apply psychological theory in the ecology of everyday life, with a focus on meaningful health-related behavior [3,4].

In this paper, we present an exploratory 28-day study in which an mHealth app delivered interventions to support people in pursuing self-selected healthy behavior change goals such as eating more vegetables, eating more slowly, and increased walking time per day. The interventions included asking people to set implementation intentions [5]—an EBI that yields medium-to-large improvements in behavior change [6]—supplemented by reminders whose delivery schedule was experimentally manipulated to explore cognitive mechanisms that might amplify the effects of implementation intentions on behavior change. The adaptive control of thought-rational (ACT-R) theory of cognition [7] is used to develop an integrated mechanistic account of the role of implementation intentions in enhancing behavior change intentions and actions; how reminders have time-varying strengthening effects on the impact of implementation interventions; and how controlled, deliberative, goal-striving behavior becomes automatic and habitual. On the basis of this account, a predictive mathematical model is fit to the data from the mHealth study.

A key driver of health care costs in the United States and other developed nations are unhealthy but changeable behaviors such as physical inactivity and eating too much, or too much of the wrong things [8]. The working assumption for our own mHealth research is that to master the complex fabric of a new healthy lifestyle, one must master and weave together a new set of healthy habits that override the old unhealthy habits. mHealth platforms are proposed as systems that can expand and intensify psychosocial and health behavior interventions beyond clinical settings into the ecology of everyday life [9] to support the formation and maintenance of desirable healthy habits.

The study and model presented in this paper are part of a larger project (called Fittle+), with several aims. First, the project involves the iterative development of an integrated pervasive computing platform for delivering and testing multiple EBIs. Second, the Fittle+ project explores how specific EBIs in the literature—such as guided enacted mastery [10] and implementation intentions—can be translated to mHealth delivery. Third, the project pursues the development of an integrated psychological model of behavior change that encompasses multiple mechanisms and addresses the dynamic effects of the EBIs and other fine-grained mHealth interventions such as reminders.

In the rest of this introduction, we present:

- A summary of the theory of planned behavior (TPB).
- A review of implementation intentions as an EBI and the hypothesized underlying cognitive mechanisms of action for implementation intentions.
- A theoretical framework for long-term behavior change based on recent cognitive neuroscience that proposes dual systems: A system of mechanisms that supports volitional goal-striving and a system that supports habit formation and execution.
- ACT-R as a mechanistic and predictive dual-system theory of goal-striving and habit formation.
- ACT-R predictions regarding the effects of different reminder schedules on the strength of memory for implementation intentions.
- A dynamical mathematical model based on ACT-R that will be fit to individuals' daily data over the course of the 28-day mHealth behavior-change study. The model addresses effects of implementation intentions, reminder schedules, and habit formation.

Theory of Planned Behavior and Self-efficacy

The path to healthy habits is not simple. Much of the focus in behavior-change theory is on the factors that initiate and strengthen the intentional goal to change and the factors that strengthen the volitional and effortful striving to translate those goal intentions into actual behavior [11,12]. Despite criticism [13], a dominant theory of behavior change is the TPB [14]. The TPB proposes that volitional behavior change is a function of the *goal intention* to perform the behavior and *perceived behavioral control*. The goal intention is in turn a function of expectancy-value beliefs and attitudes. PBC is synonymous with the concept of *self-efficacy* in social cognitive theory [15,16] and predicts that the perception of the ease or difficulty of a particular intended behavior facilitates or impedes the intention to perform the behavior.

Implementation Intentions

Goal intentions are hypothesized to be mental representations of desired behavior and end states, which are to be distinguished from *implementation intentions* that are mental representations of simple plans to translate goal intentions into behavior under specific conditions [5,17]. Interventions designed to foster the setting of implementation intentions typically ask people to specify when, where, how, and (sometimes) with whom to act on a goal intention by using if-then statements of the form: "If I encounter situation *S* then I will initiate action *A*." It is argued [18] that one reason to focus intervention efforts on implementation intentions rather than goal intentions is that medium-to-large changes in commitment to goal intentions (d=0.66) only lead to small-to-medium changes in behavior (d=0.33) [19], but implementation intentions have medium-to-large effects on goal attainment (d=0.65) [6].

Wieber et al [18] review the experimental literature and studies of physiological correlates to bolster the hypothesis that two

processes are involved in the effectiveness of implementation intentions: (1) the mental representation of situations in which the intended behavior is to take place becomes more accessible and activates the goal intention and (2) a strong associative link between a mental representation of the situation and intended behavioral action effects a heightened readiness to perform the action and the action takes less effort.

Previous research has manipulated the situation-action associative strength of implementation intentions and their effects on behavior using priming [20]. This implicates declarative memory processes [7,21] in mediating the effects of implementation intentions. This would also suggest that other ways of strengthening the declarative memory representations of implementation intentions should enhance their effectiveness—such as the explicit use of reminders to use implementation intentions. Prestwitch et al [22] presented a mHealth study that showed that SMS text message (short message service, SMS) reminders of implementation intentions promoted increased brisk walking but did not experimentally explore the specific effects of reminders.

From Volitional Goal-Striving to Habits

Habits are only gradually learned through the association of specific behaviors to triggering cues in the environment, including physical settings and previous actions. More than a hundred years of psychological research on habits suggests that there are dual systems involved in habit acquisition and strengthening [23-25]. There is (1) a deliberative or controlled goal-striving process that motivates and guides the behavior in the relevant contexts and, through repetition (2) a habit is formed that is automatically performed without effortful, controlled goal-striving. Well-practiced habits appear to be performed automatically without mediating goals, motivation, or deliberative thought (system 1) [26], but habit formation typically depends on a long period of goal-mediated, consciously controlled, exploration, repetition, and practice of behavior (system 2) [26]. A simple example of this transition is developing the habit of keeping a food diary [27]. At first, one may need to set up reminders to go through the behaviors involved in recording meals, but eventually, the behavior can become triggered somewhat automatically at the end of every meal.

Research [25] suggests that the neural circuitry directing behavior undergoes changes as habits are acquired and strengthened. As new behavior is attempted, explored, and practiced, the prefrontal cortex communicates with the striatum (basal ganglia), and the striatum communicates with the midbrain and dopaminergic mechanisms aid learning and assign value to goals. Continued practice of the behavior forms a feedback loop between the sensorimotor cortex and the striatum, creating behavioral routines that appear to be units residing in the striatum. Habit learning is consistent with the learning of other (procedural) cognitive skills [25,28].

The underlying neurological mechanisms of habit learning are consistent with computational models of reinforcement learning [25] such as temporal-difference models [29] and the Rescorla-Wagner model [7,30,31]. Learning new sequences or organizations of behavior involves learning through experience the immediate value of actions that are currently available and the estimated value of future actions and basing choices on those learned value estimations.

An ACT-R Model of Implementation Intentions, Reminders, and Habit Learning

In recent years, there has been push to develop rich, fine-grained, dynamical theories that are up to the task of predicting mHealth cause-effect relations and guiding the engineering of new personalized interventions [32]. Computational predictive models of self-efficacy (or PBC) have been developed based on dynamical control principles [33] and on cognitive theory [34]. In this paper, we extend the model of Pirolli [35] to provide a computational account of the mechanisms involved in *intention-to-behavior* processes [17] that are hypothesized to be improved by implementation intentions, reminders, and habit learning. The model presented here is based on the ACT-R theory [7], including recent extensions [36,37].

As was summarized in Pirolli [35], ACT-R [7,21] is a unified theory of how the structure and dynamics of the brain give rise to the functioning of the mind. The ACT-R simulation environment is a computational architecture that supports the development of models.

Modules and Buffers

ACT-R is composed of *modules*, processing different kinds of content, which are coordinated through a centralized *production module*. Each module corresponds to a brain region. Each module is assumed to access and deposit information into *buffers* associated with the module, and the central production module can only respond to the contents of the buffers. The *declarative memory module* stores memories of the kind of knowledge and experience that a person can attend to, reflect upon, and usually articulate in some way (eg, by declaring it verbally or by gesture). A consciously formulated implementation intention that is later remembered and acted upon is an example of something stored in the declarative memory module. The production module stores the habits and skills we display in our behavior without conscious awareness. The *goal buffer* stores and retrieves information that represents the internal intention of the system and provides local coherence to behavior. More specifically, the modules and buffers relevant to this paper include

- *Goal buffer* (dorsolateral prefrontal cortex), which keeps track of active goals and internal state of the system.
- *Production module* (basal ganglia), which matches the contents of other module buffers and coordinates their activity. The production module stores *production rules*. A production rule is a formal specification of the *flow of information* from buffers in the cortex to the basal ganglia and back again. As suggested by the literature review above, the production module is where new habits are stored. Productions have a *utility* property that is used to select the single rule that is executed.
- *Declarative module* (temporal lobe; hippocampus) and *retrieval buffer* associated with the retrieval of knowledge and past experiences from long-term declarative memory. The declarative module is where goal intentions are stored

(before they become active goals in the goal buffer) and where implementation intentions are stored.

Knowledge in the declarative module and goal buffer is represented formally in terms of *chunks* [38,39]. Chunks have *activation* levels that determine the probability and time course of chunk retrieval into a buffer. Production utilities and chunk activations are real-valued quantities produced by *subsymbolic mechanisms* in ACT-R. These subsymbolic mechanisms reflect neural-like processes that determine the time course and probability of cognitive activity and behavioral performance. The dynamics of declarative memory retrieval and production selection are determined by these subsymbolic mechanisms.

Key Components of the ACT-R Model of Intentions, Reminders, and Habits

Pirolli [35] presented an ACT-R model motivated by Tobias [40] that is modified slightly here to suit the current experiment. The model includes the following components:

- *Goal intentions*: a goal-like representation that is stored in declarative memory as a kind of prospective memory [41,42] to be turned into an active goal, in the goal buffer, in response to the right context
- *Implementation intentions:* plan-like representations that are also stored in declarative memory to be turned into concrete behaviors by production rules
- *Reminders*: SMS text messages that cue the recall of implementation intentions and thereby, increase the activation of implementation intentions through learning mechanisms so that they are more likely to be retrieved in the right context in the future
- *Habit compilation (ie, production compilation)*: repeated execution of complex sequences of cognitive and behavioral steps (multiple production rules, multiple memory retrievals) produce new, simpler production rules that require less cognitive effort the next time around
- *Utility learning*: new habits are rewarded and slowly come to dominate over the old habits

Mechanisms Underlying the Dynamics of Reminders and Memory

Figure 1 presents a subset of the ACT-R mechanisms relevant to the current model. The first two equations in Figure 1 define how the level of activation of chunks in declarative memory relates to the probability of their retrieval at any given time. The third equation for *base-level learning* defines how activation levels are increased by repeated experiences (practice) or decay with time (forgetting). These activation and base-level learning mechanisms are crucial to the ACT-R model of implementation intentions and the effects of reminders. Reminders for implementation intentions will increase their activation in declarative memory, but activation will decay as time goes by since those reminders were attended.

We hypothesize that the base-level strength of implementation intentions in declarative memory will be associated with improved success in achieving behavior-change goals. The dynamics of base-level strength will be related to frequency and timing of reminders, as well as the frequency and timing of actual use of the implementation intentions in performing behavior.

The base-level learning mechanisms defined in Figure 1 propose that each time an implementation intention is formulated, reminded, or put into practice, it receives an increment of activation (a practice effect). However, each increment of activation decays as a power function of time (the forgetting effect).

Figure 1. Some key adaptive control of thought-rational (ACT-R) subsymbolic mechanisms.

Mechanism	Equation	Description
Retrieval probability	$P_i = \frac{e^{A_i/s}}{\sum_j e^{A_j/s}}$	P_i: The probability that chunk i will be recalled A_i: Activation strength of chunk i $\sum A_j$: Activation strength of all of eligible chunks j s: Chunk activation noise
Activation	$A_i = m_n(\ldots) + \varepsilon_i$	m_n: Base-level activation reflects the recency, frequency, and schedule of use of chunk i ε_i: Random noise value
Base-level learning	$m_n(t_1 \cdots t_n) = \ln\left(\sum_{i=1}^{n}(t_n - t_i)^{-d_i}\right)$ $d_i = c\exp(m_{i-1}) + a$	n: The number of experiences for chunk t_i: The time of the i^{th} presentation d_i: The decay rate for the i^{th} presentation a, c: Scale and slope parameters
Utility learning	$U_i(n) = U_i(n-1) + \alpha[R_i(n) - U_i(n-1)]$	$U_i(n-1)$: Utility of production i after its $n\text{-}1^{st}$ application $R_i(n)$: Reward production receives for its n^{th} application $U_i(n)$: Utility of production i after its n^{th} application
	$P_i = \frac{e^{U_i/s}}{\sum_j e^{U_j/s}}$	P_i: Probability that production i will be selected U_i: Expected utility of the production determined by the utility equation above U_j: is the expected utility of the competing productions j

The rate of decay of each increment of activation depends on the strength of activation at the time of the reminding or practice: at longer intervals between remindings or practice, the activation levels are lower, and subsequent forgetting occurs less quickly (the spacing effect). When reminding or practice is spaced closely, the forgetting occurs more quickly.

Habit Learning Mechanisms

The last two equations in Figure 1 define utility learning and the relation of utility to the probabilistic choice of production rules to execute. These utility mechanisms are crucial to the ACT-R model of habit formation.

Also important in the ACT-R model of habit formation is the mechanism of *production compilation* [7,43,44] by which new production rules are acquired. A new production rule is generated every time two production rules are executed in sequence. The mechanism works to create new rules that eliminate internal cognitive processing, such as the need to retrieve information from the declarative module or set and maintain sequences of goals. Production compilation is viewed as the mechanism underlying the acquisition of new habits. Utility learning is a variety of reinforcement learning similar to temporal-difference learning [29] and Rescorla-Wagner learning [31]. According to the utility learning equation, the utility of a new production rule (habit) is gradually adjusted until it matches the average reward for using the production.

ACT-R Predictions About Reminding Schedules and the Memory Strength of Implementation Intentions

Figure 2 presents the reminding schedules used in our 28-day study. The experiment manipulated the total frequency of reminders (7 or 14 reminders over 28 days) and the distribution of presentation (massed or distributed). Each vertical bar in Figure 2 indicates the day on which a reminder was sent to our participants. In the massed conditions, some reminders occur with less temporal spacing. The base-level learning parameters used to plot the base-level activation in Figure 3 are from Pavlik and Anderson [37], and they illustrate the practice, forgetting, and spacing phenomena: reminders are expected to boost up the base-level activation, but the activation decays without further practice, and distributed reminders are forgotten less quickly.

Each plot in Figure 3 also presents the predicted mean activation level of the implementation intention over the full 28 days for each condition (upper left corner of each plot). Note that at low frequency of reminders, the mean activation level in the massed condition is greater than that of the distributed condition, but at high frequency, the mean activation of the distributed condition is greater than the massed condition. Thus, there is a predicted interaction of reminder distribution (massed, distributed) by frequency (low, high) and specifically, the average activation levels for the implementation interventions are predicted to be high frequency-distributed > high frequency-massed > low frequency-massed > low frequency-distributed. Behavior-change data from the mHealth experiment will be used to test for this predicted interaction and the specific ordering of success rates predicted by the model.

Figure 2. Reminder schedules used in the experiment.

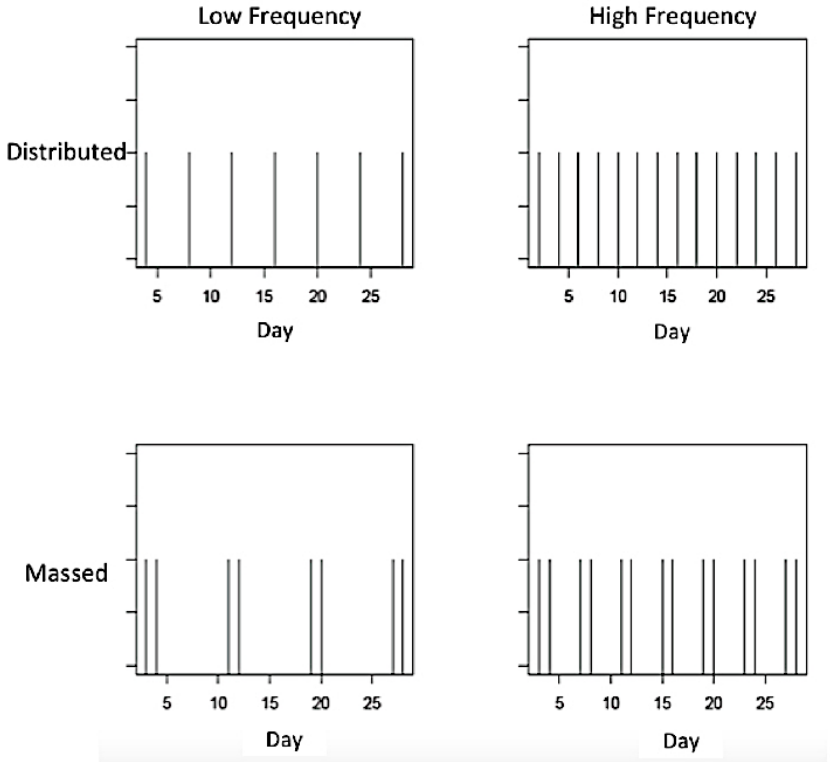


Figure 3. Simulated base-level learning of implementation intentions as a function of different reminder schedules.

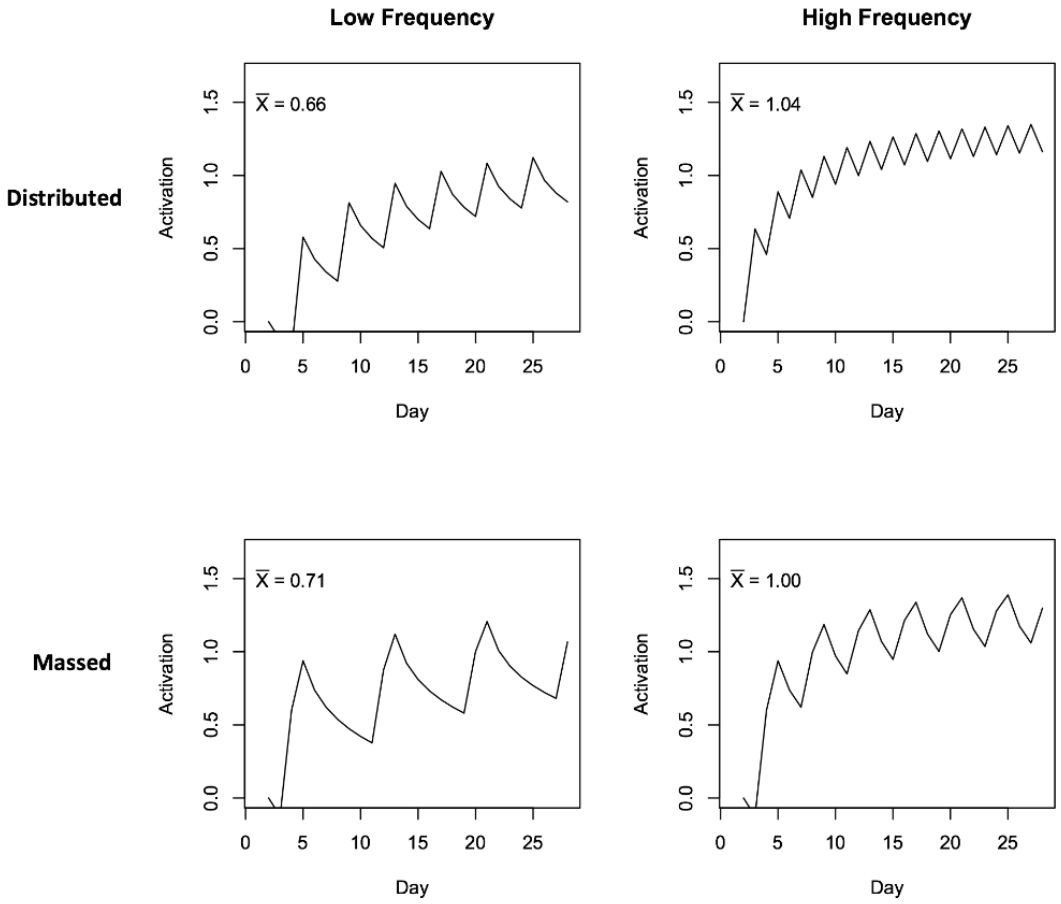
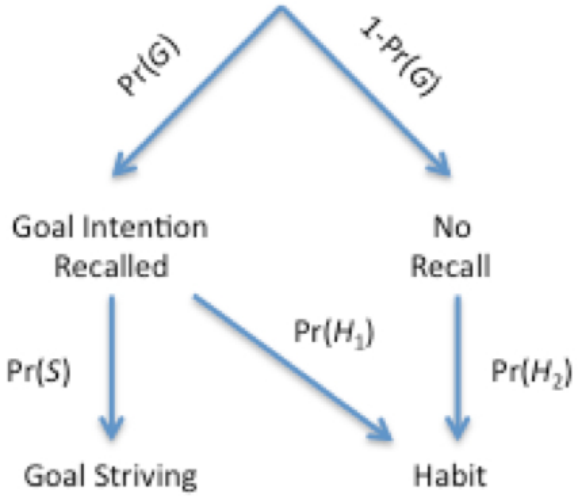


Figure 4. Probability tree for cognitive states and processes in the model.



A Dynamical Model of Reminding and Habit Learning Effects for Each Individual

As discussed previously, theory suggests that behavior change involves dual systems: a goal-striving system that is heavily dependent on declarative memory and a habit-forming system that acquires and reinforces more automatic procedures for performance. Figure 4 captures the dual-system model that was fit to the data from the mHealth experiment. The model presented here is consistent with the ACT-R model presented above but is also similar to that of Pavel et al [45]. It is a dual-system model that includes (1) a habit or reinforcement learning system and (2) a declarative memory system with base-level learning of memory activation. The declarative memory model includes the Pavlik and Anderson [37] mechanisms to account for effects that may occur when remindings happen at variable spacing, which were manipulated in the experiment. If a reminder happens, there is an increase in the activation of the intention memory. If a goal success happens, there is habit reinforcement, as well as a strengthening

of the activation of declarative memories for goal intentions and implementation intention memories.

Figure 5 presents the details of the model in terms of 12 equations. On a given day, in the appropriate context for doing their goal behavior, a participant may recall their goal intention or not with probability Pr(G). If the goal intention is recalled, there are two routes to successful behavior: the goal behavior may be achieved through effortful goal striving with probability Pr(S), or if a habit has formed, it can be achieved by the new habit routines with probability Pr(H_1). At first, goal-striving will be more probable, and after many repetitions, the habitual behavior will be more probable. After the habit is well practiced, even if the goal intention is not explicitly recalled, it may be executed with probability Pr(H_2).

Altogether, the probability of participant success on a given day will be given by equation 1 (Figure 5). We can rewrite this as a dynamical equation dependent on day t as equation 2 (Figure 5). Where *Success* (t) is the probability of success of performing the goal, G (t) is the probability of remembering the goal intention, $H_1(t)$ is the probability of a habit routine given the goal intention has been recalled, and $H_2(t)$ is the probability of the habit given no goal intention recall.

The probability of recall based on activation level is given in equation 3 (Figure 5), where A_{II} is the base-level activation from reminders, and A_A is the base level activation from past experiences of actually doing the behavior (declarative memory of the experiences), and β_0, β_1, and β_2, are scaling and slope parameters to be estimated.

As shown in Figure 1, ACT-R chooses productions based on the utility learned for those productions. The choice of goal-striving production versus habit productions is given by Equation 4 (Figure 5), where U_θ is a threshold utility that essentially captures the other behaviors competing with (or impeding) the choice of the goal behaviors. U_s is the utility of the goal-striving productions that we assume is also dependent on the activation strength of the implementation intention in memory as given in equation 5 (Figure 5).

The habits associated with the goal behavior are learned according to the utility learning equation 6 (Figure 5), where R is the reward value associated with successfully performing the targeted goal behavior. The performance of a habit competing with goal-striving is given by the probability $H_1(t)$ in equation 7 (Figure 5). In the case where the goal intention has not been recalled, but there is a habit that is being learned, the habit productions just compete with the background threshold, and the probability of the habit is $H_2(t)$ in equation 8 (Figure 5).

What remains to be defined is the declarative memory model that captures the base-level learning effects from the reminders and the successful behavior experience. As presented in Figure 1, each time, t_i, a reminder, or experience happens, there is an increment of activation that decays as a power function with decay parameter d_i. The total base-level activation is just the log of the sum of all those decaying increments as given in equation 9 (Figure 5). One complexity is that the decay parameter on each activation increment for a reminder or experience can vary as a function of the current level of activation, as defined in equation 10 (Figure 5), where c and a are scaling and slope parameters. This fits the observation that forgetting is slower when interpractice time is longer (spaced). So the base-level learning for the implementation intention, in equation 11, is defined by the times at which the reminder happened ($r_1^+, \ldots r_k^+$), and the base-level activation because of successful behaviors given in equation 12 is defined by the times at which the experiences happened, ($g_1^+, \ldots, g_k^+$).

Summary of Aims

A 28-day exploratory mHealth experiment was conducted to investigate ACT-R predictions about the effects on achievement of behavior-change goals of implementation intention reminders and of prior goal achievement. The experiment uses an mHealth app that engages users to select goals to do new healthy behaviors (eg, eating more vegetables, eating more slowly, and increased walking time per day) to be performed every day and to set implementation intentions to do those behaviors.

The aims of this study were to

- Perform exploratory data analyses for signature phenomena predicted by the ACT-R theory. Memory for implementation intentions is predicted to (1) improve with the cumulative frequency of reminders and cumulative frequency of performance of the goal behaviors (practice effects) and (2) diminish with the time since presentation of past reminders and time since past goal performances (forgetting effects).
- Test-specific ACT-R predictions about the effects of reminder schedules on memory for implementation intentions, as revealed in participants' rates of daily adherence to behavior-change goals (summarized in Figures 2 and 3)
- Model each individual's daily goal achievement data using a dynamical mathematical model based on ACT-R. The ACT-R model captures the mechanisms underlying the role of reminders in amplifying the effects of implementation intentions during the volitional goal-striving phase of behavior change, as well as the gradual learning of new habits.

Figure 5. The adaptive control of thought-rational (ACT-R) dual-system model.

Number	Equation	Summary
(1)	$Pr\,(Success) = Pr\,(G)\,[Pr\,(S) + Pr\,(H_1)\,] + [1 - Pr\,(G)\,]Pr\,(H_2)$	Probability of goal success
(2)	$Success(t) = G(t)[S(t) + H_1(t)] + [1 - G(t)]H_2(t)$	Goal success dynamics
(3)	$G(t) = \frac{exp\,(\beta_0+\beta_1 A_{II}(t)+\beta_2 A_A(t))}{1+exp\,(\beta_0+\beta_1 A_{II}(t)+\beta_2 A_A(t))}$	Probability of recall based on activation level
(4)	$S(t) = \frac{exp\,(U_S(t))}{exp(U_\theta(t))+exp(U_S(t))+exp(U_h(t))}$	Choice probability for goal-striving
(5)	$U_S(t) = \beta_3 + \beta_4 A_{II}(t) + \beta_5 A_A(t),$	Utility of goal-striving
(6)	$U_H(t) = U_H(t-1) + \alpha[R - U_H(t-1)]$	Utility of performing habit
(7)	$H_1(t) = \frac{\exp\,(U_H(t))}{\exp(U_\theta(t))+\exp(U_S(t))+\exp\,(U_H(t))}$	Choice probability of habit competing with goal-striving
(8)	$H_2(t) = \frac{\exp\,(U_H(t))}{\exp(U_\theta(t))+\exp\,(U_H(t))}$	Choice probability of habit when no goal recalled
(9)	$m_n(t_1 \cdots t_n) = \ln\,(\sum_{i=1}^{n}(t_n - t_i)^{-d_i})$	Total base-level activation for reminders and experiences
(10)	$d_i = c\exp(m_{i-1}) + a$	Memory decay as a function of current activation
(11)	$A_{II}(t) = m_n(r_1^+ \cdots r_k^+)$	Base-level learning for an implementation intention as a function of reminders
(12)	$A_A(t) = m_n(g_1^+ \cdots g_k^+)$	Base-level learning as a function of successful behavior performance

Methods

We first present the mHealth experiment and then summarize the analysis methods that were applied to explore the ACT-R model of implementation intentions, remindings, and habit formation.

Experiment

Design

An incomplete factorial design was used in this experiment (see Table 1). All participants were asked to choose a healthy behavior goal and to set implementation intentions. Participants were stratified by *self-efficacy*: their confidence to complete the selected behavior goal. Participants were assigned to one of two *reminder* conditions: presented versus absent. Self-efficacy and reminder conditions were crossed. Nested within the reminders presented condition was a 2 X 2 crossing of *frequency* of reminders sent (high, low) by *distribution* of reminders sent (distributed, massed). Participants in the low frequency condition were sent a total of 7 reminders, and participants in the high frequency condition were sent 14 reminders. Participants in the distributed conditions were presented reminders at uniform intervals: high frequency distributed participants were sent reminders on days 2, 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, whereas low frequency distributed participants were sent reminders on days 4, 8, 12, 16, 20, 24, and 28. Participants in the massed distribution conditions were sent reminders in clusters (lower row of plots in Figure 2): high frequency massed participants were sent reminders on days 3, 4, 7, 8, 11, 12, 15, 16, 19, 20, 23, 24, 27, and 28 and low frequency massed on days 3, 4, 11, 12, 19, 20, 27, and 28. In total, there were 10

conditions: 2 (self-efficacy) X 2 (frequency) X 2 (distribution)=8 in the reminder-present condition, plus 2 (self-efficacy) in the reminder-absent condition.

Participants

N=64 participants were recruited using the email list of a large university, Craigslist, and Nextdoor. Participants were paid US $50 in gift cards for their participations. Participants ranged in age from 25 to 71 years, with a median of 30 years. Internet protocol addresses suggested that participants came from eight different US states, and one came from India, with the majority coming from California and Michigan. Participants were randomly assigned to the 10 cells of the incomplete factorial design, resulting in slightly unbalanced cells with 5 to 8 participants per cell (Table 1). Pooling across the cells in Table 1, there were N=34 participants in the low self-efficacy conditions and N=30 in the high self-efficacy conditions. Pooling across reminders conditions, there were N=51 participants in the reminders presented conditions and N=15 in the reminders absent conditions. There were N=23 low frequency and N=28 high frequency participants and N=26 distributed distribution and N=25 massed distribution participants.

Materials

Four types of health behavior goals (habits) were developed by the nutrition and exercise specialist on our team: eating slowly (12 habits), walking (19 habits), food journaling (11 habits), and eating vegetables (6 habits). N=1500 participants from Mechanical Turk rated subsets of these habits on habit *difficulty* and perceived *self-efficacy*. The difficulty question was " *how difficult is it for you to complete this goal everyday for the next 7 days?:* 1-10 scale," and the self-efficacy question was " *how confident are you that you can complete this goal everyday for the next 7 days?: 0%-100% scale.* " Reported difficulty and self-efficacy rating were highly negatively correlated. Food journaling tasks did not show much variability in difficulty ratings and were dropped from the study. For the remaining three habits: eating slowly, walking, and eating vegetables, we selected the three most difficult habits and the three easiest habits. An example habit was "Stretch for 10 minutes and walk for 30 minutes in the afternoon." An example of an implementation intention reminder for this habit was "Remember to stretch for 10 minutes and walk for 30 minutes in the afternoon—at in my neighborhood with a friend." The complete list of habits is in Multimedia Appendix 1.

PARC Coach

PARC Coach is a mobile app developed to study behavior change interventions in an mHealth setting. It implements only the most central features to reporting behavior and delivering interventions. PARC Coach has a reporting home page on which people report whether they have met their daily goal. Informational pages were available for every behavioral goal.

Procedure

Upon creating an account and first time logging in, they were asked to select a class of habit to pursue: eat slowly, walking, or eating more vegetables. Upon habit selection, participants were randomly assigned to either a high self-efficacy or low self-efficacy goal for that habit class. Upon being assigned the specific habit, participants were asked to rate their self-efficacy for achieving the assigned goal with the question "*how difficult is it for you to complete this goal everyday for the next 7 days?:* 1-10 scale."

Self-reported daily goal achievement reports were collected in the PARC Coach app. The app contains a reporting page in which participants click a button to indicate whether they did their goal behavior.

Participants were then asked to set an implementation intention with the following components (possible responses in parentheses): (1) *Which part of day will you do practice this habit?* (morning, afternoon, evening) or *which meal will you like to try this at?* (breakfast, lunch, or dinner), (2) *Where will you do this activity?*, (3) *Who will you do this activity with?*, and (d) *When will you do this activity?* (event time).

If participants were in a reminder condition, they were additionally asked *How long before the event would you like to be reminded of your task?* (reminder duration). In the reminder condition, the participants' selection of an event time and a reminder duration were combined with the reminder schedule to compute when to send the reminder.

Table 1. Number of participants assigned to cells of the incomplete factorial design.

Self-efficacy	Reminders	Frequency	Distribution	Participants (n)
Low	Presented	Low	Distributed	6
Low	Presented	Low	Massed	6
Low	Presented	High	Distributed	7
Low	Presented	High	Massed	8
Low	Absent			7
High	Presented	Low	Distributed	6
High	Presented	Low	Massed	5
High	Presented	High	Distributed	7
High	Presented	High	Massed	6
High	Absent			6

Reminders were sent by SMS to the participants' mobile phone. The content of the reminders was determined by the habit and the participants' implementation intentions. Participants were expected to acknowledge reminders by clicking "OK." The reminder was canceled if it wasn't acknowledged until the event time. The sending of reminders and the acknowledgment of their receipt was automatically logged by the PARC Coach app.

Analyses

Exploratory Data Analyses for Signature Memory Phenomena

ACT-R predicts that the effects of reminders on the memory activation levels of implementation intentions will exhibit practice and forgetting effects. Memory activation is predicted to increase with the frequency with which reminders are processed (practice) and decrease as with decreases in the recency since reminders were processed (forgetting). When a reminder is sent to participants, there is a chance that they may ignore the reminder. When participants acknowledged the receipt of reminders, we assume that is an indicator that they actually attended to and processed the reminder.

We performed a set of exploratory analyses on the basic relationship of reminder schedules to achieving behavior-change goals (adherence). For every participant, on every day, the dependent variable of goal adherence was coded (success=1, failure=0). For every participant, on every day, we also coded how many times a reminder had been acknowledged as received since the start of the experiment (*frequency acknowledged*) and how many days since the last reminder had been acknowledged (*recency acknowledged*); how many time a reminder has been sent since the beginning of the experiment (*frequency sent*) and how may days since the last reminder had been sent (*recency sent*); and how many days adherence had been reported in the past (*frequency adherence*) and how may days since the last report of success (*recency adherence*).

We, for each type of input variable (reminder acknowledged, reminder sent, past goal adherence) at every level of frequency and recency, computed the mean probability of adherence (ranging from 0-1) and computed simple linear regressions of the form

$$Adherence \sim \beta_{F0} + \beta_{F1*} Frequency,$$

and

$$Adherence \sim \beta_{R0} + \beta_{R1*} Recency,$$

and computed the goodness-of-fit R^2 statistics. In addition, we analyzed the contribution of self-efficacy and reminder frequency and recency with a logistic regression:

$$logit(Adherence) \sim \beta_{0i} + \beta_0 + \beta_1 S + \beta_3 R$$

where *S*=self-efficacy (a categorical variable), *F*=frequency acknowledged, *R*=recency acknowledged, and is a random coefficient estimated for each participant *i*.

Analysis of Specific ACT-R Predictions About Reminder Schedules

We performed a logistic regression analysis of goal adherence data within the 2 (frequency) X 2 (distribution) factorial conditions in which reminders were presented. This analysis serves as a test for the specific pattern of a priori ACT-R predictions about the effects of reminder schedules on memory for implementation intentions. Those predictions are summarized in Figures 2 and 3. Specifically, based on the mean base-level activations, ACT-R predicts an interaction of frequency X distribution with the average activation levels for the implementation interventions predicted to be high frequency-distributed > high frequency-massed > low frequency-massed > low frequency-distributed.

Fit of the ACT-R Dual-System Mathematical Model to Individual-Level Data

A fit of the ACT-R mathematical model (defined in equations 1-12) to the data for every individual and every day was obtained by minimizing the Brier score between model-predicted probability success and observed success using the *R* optimx package using a quasi-Newton method called limited-memory BFSG, which allows one to bound the parameter search (by providing upper and lower boundaries).

Results

Signature Practice and Forgetting Phenomena Because of Reminders and Performance

Practice and forgetting effects are key signatures of declarative and procedural memory [46,47]: improvements generally accrue with repeated practice or remindings and decay over time without continued practice or reminding. Figure 6 demonstrates that these signature phenomena are apparent in the success rates of the participants. Each point in the plots is the mean observed probability of participants reporting success at achieving a goal on a given day as a function of different frequency and recency factors. The top two plots present the probability of reported success as a function of frequency and recency of past goal success. The middle two plots present the observed probability of success as a function of the frequency and recency of sent reminders. The bottom two plots present the observed probability of success as a function of frequency and recency of acknowledging the sent reminders. Each plot also presents a best-fit linear regression line, as well as the adjusted R^2 for the regression.

In general, the frequency and recency of past success (adherence) tends to show medium-to-strong relationships to current success. The frequency and recency of acknowledged reminders tends to show stronger effects than the frequency and recency with which those reminders were sent. This is unsurprising, as one would expect that sent reminders could be ignored, but acknowledgment indicates that the participant actually processed the reminder.

Table 2 presents a logistic regression, with the daily success or failure as the response variable, participants as a random effect, and the predictors being the self-efficacy factor and the observed frequency and recency of reminder acknowledgment. Self-efficacy was not significant, frequency acknowledged was marginally significant, and the recency acknowledged was highly significant.

Figure 6. Factors related to the likelihood of a participant succeeding at behavioral goal on a given day. Frequency of adherence is the cumulative number of past "success report" days. Recency of adherence is the number of days since last "success report." Frequency sent is the cumulative count of reminders previously sent. Recency sent is the number of days since the last reminder was sent. Frequency acknowledged is the cumulative count of previously acknowledged reminders. Recency acknowledged is the numbers of days since the last acknowledgement of a reminder. Adjusted R^2 values are based on linear regressions.

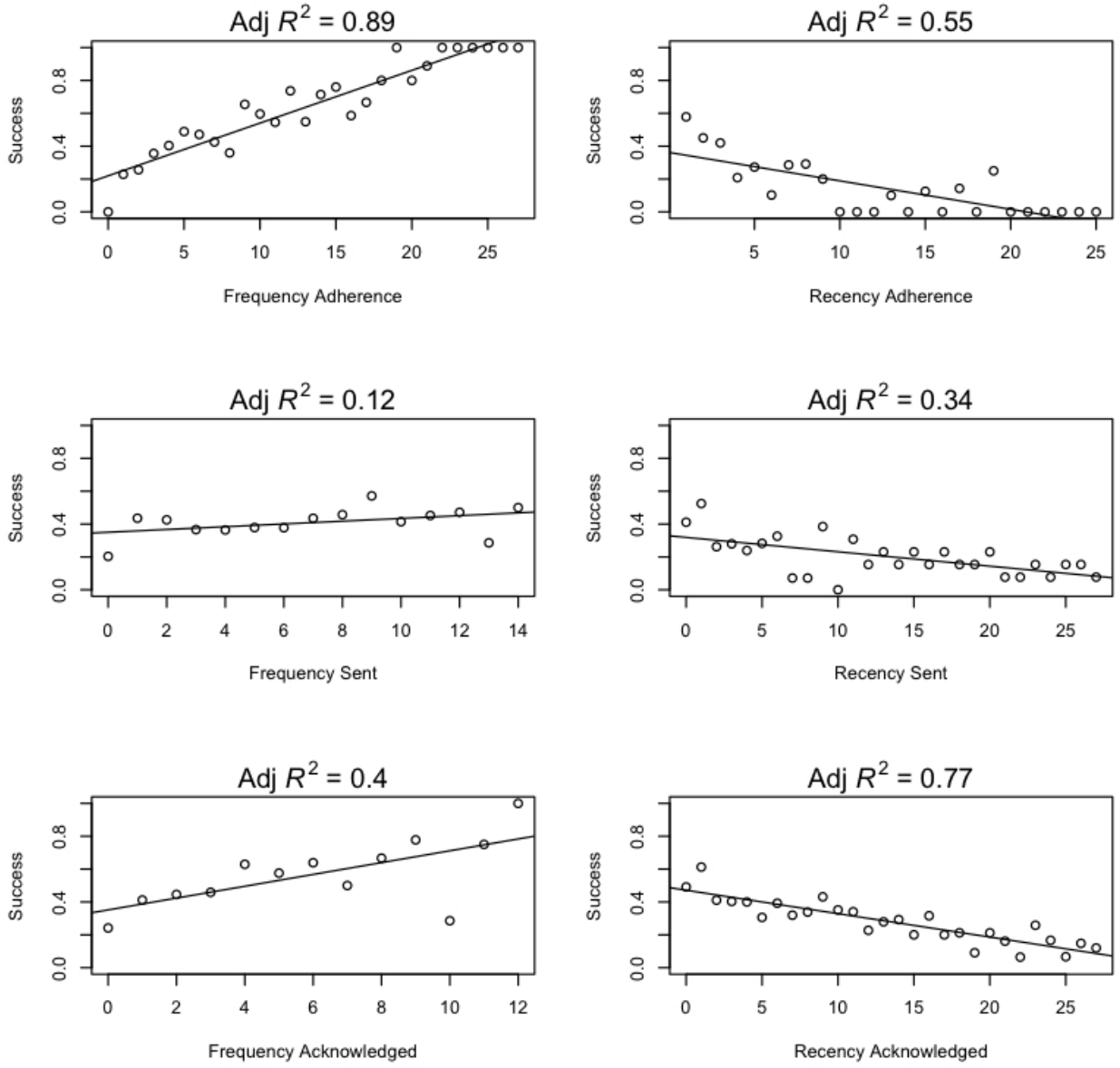


Table 2. Logistic regression of daily success in achieving self-selected goals on self-efficacy and frequency and recency of acknowledged implementation intention reminders.

Predictor	Coefficient (standard error)	Odds ratio (95% CI)	*P* value
Intercept	−0.5696 (0.3185)	0.5657 (−1.2231 to 0.0602)	.007
Low self-efficacy	−0.1197 (0.4180)	0.8872 (−0.9655 to 0.7232)	.77
Frequency acknowledged	0.0694 (0.0410)	1.0717 (−0.0116 to 0.1505)	.09
Recency acknowledged	−0.0490 (0.0104)	0.9522 (−0.0700 to −0.2852)	<.001

An analysis of deviance also showed that including frequency acknowledged to the base model of recency acknowledged was marginally significant χ^2_1=2.8 *P*=.09, and adding self-efficacy to the acknowledgment factors was not, χ^2_1=0.1.

Effects of Implementation Intention Reminder Schedules on Behavior Success: A Test of ACT-R Memory Predictions

Table 3 shows the mean proportion of days on which participants reported succeeding at their behavior change goals.

Notably, the condition in which participants received reminders was significantly better than the conditions in which others did not receive reminders, (logistic regression on reminders vs no reminders: coefficient=2.018, SE=0.572, OR=7.52, 95% CI 0.9037-3.2594, *P*<.001).

The pattern of success rates in Table 3 suggests an interaction of reminder frequency by distribution, as predicted by the ACT-R base-level learning theory (Figure 3). A logistic regression was performed on the average rate of participant success in performing their goals over 28 days in the 2 X 2 factorial conditions of distribution (distributed, massed) X frequency (low, high). The results of this regression are presented in Table 4. Although there are no main effects of the frequency or distribution variables, there was a highly significant interaction. An analysis of deviance showed that the model with the interaction term was significantly better than a reduced model without the interaction term, χ^2_1=13.056, *P*=.0003. Post hoc linear contrasts show that the high frequency-distributed condition produced higher participant success than the high frequency-massed, *z*=4.441, SE=0.2853, *P*<.001; the high frequency-massed produced higher success than the low frequency-massed, *z*=3.041, SE=0.2609, *P*=.003; and the low frequency-massed produced marginally higher success than the low frequency-distributed, *z*=1.929, SE=0.2663, *P*=.069.

Table 1 shows that each of the 2 X 2 cells of the frequency X distribution factorial have N=11 to N=14 participants. Gelman and Carlin [48] suggest that such small-sample experiments warrant an analysis of two kinds of potential experimental design errors: (1) the probability that the estimate of the effects is in the wrong direction (*Type S [sign] error*) and (2) the factor by which the magnitude of an effect might be overestimated (*Type M [magnitude] error*). We followed the procedure recommended by Gelman and Carlin. First, we went to the literature to identify a possible range of true effects sizes for our experiment. A recent meta-analysis [49] of the effect size (ES) of behavior-change interventions [50] indicates an ES=0.37 in improving physical activity and diet over the short term (<12 weeks), with a 95% CI 0.26-0.48. Next, using the *retrodesign()* function [48] we determined the *power*, *Type S*, and *Type M* error for the interaction effect in Table 4. At ES=0.37, the probability that a replication would be statistically significant at alpha=.05 was *power*=0.39, the probability that a replicated estimate would have the incorrect sign was *Type S error*=0.0003, and the expectation of the ratio of the estimated interaction effect to a true ES=0.37 was *Type M error*=1.56. It is unlikely that the sign of the significant interaction is incorrect (*Type S error*), and the magnitude of the interaction in Table 4 is likely overestimated by a factor of 1.56 (*Type M error*). For ES=0.26: *power*=0.22, *Type S error*=0.0038, *Type M error*=2.13; for ES=0.48: *power*=0.59, *Type S error*<0.0001, *Type M error*=1.30.

These analyses suggest that the pattern of participant success at their behavior-change goals over 28 days is consistent with the ACT-R theoretical predictions of how the variations in reminder schedules affect the base-level activation of participants' implementation intentions in declarative memory.

Fit of the ACT-R Dual-System Mathematical Model to Individual-Level Data

Figure 7 plots the goal success predictions of the dynamical model against the observed data as functions of past adherence (frequency and recency) and reminders acknowledged (frequency and recency). The points are the observed probabilities of success, and the lines are the model predicted probabilities. Each point is the mean of the observed individual daily success for participants at a given level of recency or frequency, and similarly, the lines are the means of the model's predictions for each individual on each day, pooled by level of frequency and recency. Parameter estimates for the fitted more are presented in Table 5. The Brier score on this fit was 0.1724.

Figure 7 suggests that the model is doing a reasonable job of predicting the observed frequency and recency effects because of the reminding interventions, as well as the recency and frequency effects because of the practice of the target behavior. The model does appear to predict a more sublinear relationship between the recency of reminder acknowledgment than is present in the observed data.

The parameters estimated in Table 5 also appear to be generally reasonable. In predicting the probability of recalling a goal and striving to do a new behavior, the activation levels of implementation intentions and memory for past goal performance are both associated with nonzero positive weights (β_1 and β_2). Similarly, the base-level activation of implementation intentions and memory for past goal performance are also positively weighted in determining the utility of goal-striving performance (β_4 and β_5). The values of parameters, *a*, *c*, determining the base-level learning that we estimated are different than those found in controlled laboratory studies [37], where alpha=.177 and *c*=0.217. It should be noted that the Pavlik and Anderson study involved multiple blocks of multiple trials within a study day, and the spacing of reminders was manipulated within a 60 to 90 min study session. In our study, we were manipulating the spacing of reminders over many days.

Table 3. Mean proportion of days on which participants reported success in achieving their behavior-change goal (standard deviation in parentheses).

Distribution	Frequency	
	Low	High
Distributed, mean (SD[a])	0.32 (0.33)	0.55 (0.24)
Massed, mean (SD)	0.34 (0.22)	0.38 (0.26)
No reminders, mean (SD)	0.18 (0.23)	

[a]SD: standard deviation.

Table 4. Logistic regression on the average rate of participant success in performing their goals over 28 days in the 2 X 2 factorial conditions of distribution (distributed, massed) X frequency (low, high).

Predictor	Coefficient (standard error)	Odds ratio (95% CI)	*P* value
Intercept	−0.6305 (0.1197)	0.5323 (−0.8681 to 0.3984)	<.001
High frequency	0.1630 (0.1584)	1.1770 (−0.1467 to 0.4746)	.30
Distributed distribution	−0.1167 (0.1672)	1.1238 (−0.4448 to 0.2112)	.49
High frequency X distributed	0.7994 (0.2215)	2.2242 (0.3656-1.2341)	<.001

Figure 7. Fit of the adaptive control of thought-rational (ACT-R) dual-system model to daily success in performing behavior goals.

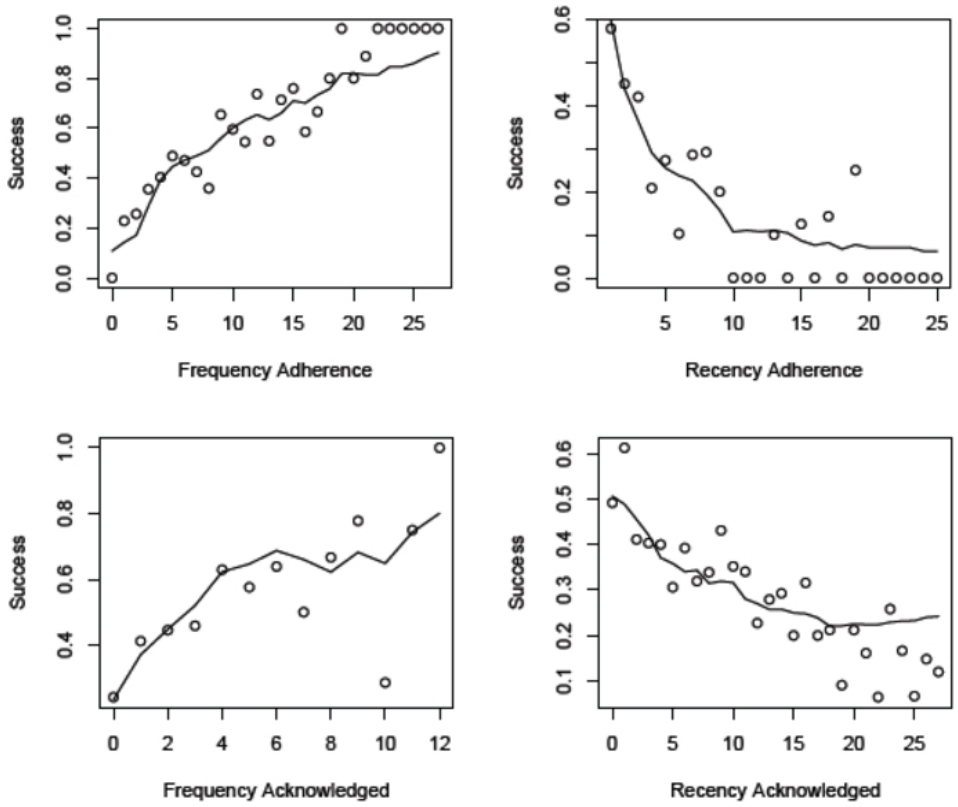


Table 5. Parameter estimates for the adaptive control of thought-rational (ACT-R) dual-system model.

Parameter	Value	Description
β_0	8.107708	Scaling parameter on activation for predicting goal recall
β_1	4.896597	Weight, implementation intention activation in predicting probability goal recall
β_2	3.535064	Weight, memory activation of performing goals in predicting probability goal recall
β_3	−0.732805	Scaling parameter on utility of goal striving productions
β_4	0.297554	Weight, implementation intention activation in utility of goal striving productions
β_5	1.396243	Weight, memory activation of performing goals in utility of goal striving productions
a	1.000000	Scaling parameter on base-level activation learning
c	0.077193	Slope parameter on base-level activation learning
U_0	−3.818326	Initial utility of new habit
α	0.291842	Utility learning rate for the habit
R	0.000000	Reward value for new habit

Discussion

Consistent with previous research on implementation intentions [22], we found that reminders of implementation intentions sent as SMS text messages have a boosting effect on success in achieving behavior change goals. Further exploratory analysis suggested that the effects of reminding on behavioral success appear to show signatures long associated with human memory: a decay of effectiveness because of forgetting and improvement with repetition. The analysis also indicates improvement with behavior practice, which is ubiquitous in the procedural learning of skills and habits. A mathematical model based on ACT-R captures the dynamics of people accomplishing their goal behaviors under the influence of implementation intentions, reminders, and their own past performance.

For practitioners, prior research had suggested the utility of using implementation intentions in mHealth [22,51]. Our results suggest that reminders can boost the effectiveness of implementation intentions in ways predicted from basic memory theory. Our results also suggest the importance of the recency factor: the number of days since a reminder was acknowledged accounts for a large proportion of the variance in goal adherence (Figure 6) and was highly significant in the analysis reported in Table 2. Attending to reminders is associated with performing the behavior soon thereafter. This might suggest that mHealth reminders be triggered by some combination of nonadherence and time since the last reminder was sent. One concern might be that sending more reminders might just cause users to decide to ignore them. As a post hoc analysis, we examined whether the acknowledgment of reminders was reduced when reminders were temporally clustered in the massed condition, in contrast to the distributed condition. We found no statistical difference, but we still were not sending reminders more than once a day.

It has been argued [52] that the current menagerie of behavior change theories [11,50] needs to be refined or replaced with precise models that yield predictions at the granularity of assessments and interventions that are delivered by mHealth systems in the ecology of everyday life. Such models would provide a rigorous foundation for engineering sophisticated, individualized interventions that are optimally delivered in the right contexts at the right time. In the present research, we worked from the existing ACT-R theory [7] to propose a model of goal-striving and habit formation that would predict the effects of implementation intentions and reminders. Previous research [18] had suggested that human declarative memory mechanisms were implicated in the effectiveness of this EBI. ACT-R predicts the dynamics of declarative memory retrieval in response to additional training or use, and so the theory was extended to make predictions about the dynamics of implementation intention effectiveness as a function of the timing of reminder interventions.

Related ACT-R models [34] have been developed to predict the dynamics of self-efficacy and goal success in users of an mHealth app called DStress [10]. In that case, there were two kinds of modeling: (1) full simulations in the ACT-R cognitive architecture and (2) dynamical mathematical models approximating the ACT-R mechanisms, similar to the approach presented here. The ACT-R–inspired mathematical models are also similar to the dynamical models presented in Pavel et al [45], which also used a dual-systems approach for goal-striving and habit formation. In other applied domains such as cognitive tutoring [53] and language learning [36], it has been useful to develop user models that are approximations to the detailed ACT-R simulation architecture, yet, still support prediction by computation (see also, [54]).

To repeat an argument made by Pirolli [34], the motivation for developing mHealth theories by extending theories of the human cognitive architectures rests on four theses [55,56]: (1) the *integration thesis*, that cognitive architectures provide a unified account of how the modules of the mind function together to produce coherent behavior and can provide a basis for an integration across specialized domains of EBIs, theories of behavior change, and multiple systems and mechanisms of action in behavior change; (2) the *decomposition thesis*, that long-term behavior change can be decomposed to learning and intervention events occurring at a much finer granularity of time; (3) the *modeling thesis*, that models in cognitive architectures can provide a basis for bridging those events at the small scale to the dynamics of behavior change occurring at the large scale; and (4) the *relevance thesis*, that long-term changes and outcomes can be improved by modeling and predicting specific just-in-time interventions in contexts that are occurring at the smaller time scales in the ecology of the everyday lives of people wishing to change.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1346066 to PP and Michael Youngblood. Correspondence should be addressed to PP, 2958 Sloat Blvd, San Francisco, CA 94116. The data in this original article were collected in accordance with a protocol approved by the PARC Human Subjects Committee, which is an institutional review board with Federalwide Assurance from the US Department of Health and Human Services.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Habits used in the Implementation Intention Study.

[PDF File (Adobe PDF File), 20KB - jmir_v19i11e397_app1.pdf]

References

1. Silva BM, Rodrigues JJ, de la Torre Diez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003]
2. Rotheram-Borus MJ, Swendeman D, Chorpita BF. Disruptive innovations for designing and diffusing evidence-based interventions. Am Psychol 2012 Sep;67(6):463-476 [FREE Full text] [doi: 10.1037/a0028180] [Medline: 22545596]
3. Baumeister RF, Vohs KD, Funder DC. Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? Perspect Psychol Sci 2007 Dec;2(4):396-403. [doi: 10.1111/j.1745-6916.2007.00051.x] [Medline: 26151975]
4. 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]
5. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493-503. [doi: 10.1037/0003-066X.54.7.493]
6. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: a meta-analysis of effects and processes. Adv Exp Soc Psychol 2006;38:69-119. [doi: 10.1016/S0065-2601(06)38002-1]
7. Anderson JR. How Can the Human Mind Occur in the Physical Universe?. Oxford, UK: Oxford University Press; 2007.
8. Thorpe KE. Nccor. 2009. The future costs of obesity: national and state estimates of the impact of obesity on direct health care expenses URL: http://www.nccor.org/downloads/CostofObesityReport-FINAL.pdf [accessed 2017-11-04] [WebCite Cache ID 6ujojS3Pq]
9. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
10. Konrad A, Bellotti V, Crenshaw N, Tucker S, Nelson L, Du H, et al. Finding the adaptive sweet spot: balancing compliance and achievement in automated stress reduction. New York, NY, USA: ACM Press; 2015 Presented at: CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul, Republic of Korea p. 3829-3838. [doi: 10.1145/2702123.2702512]
11. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of Behavior Change Theory. Great Britain: Silverback Publishing; 2014.
12. Rimer BK, Brewer NT. Introduction to health behavior theories that focus on individuals. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2008.
13. Sniehotta FF, Presseau J, Araújo-Soares V. Time to retire the theory of planned behaviour. Health Psychol Rev 2014;8(1):1-7. [doi: 10.1080/17437199.2013.869710] [Medline: 25053004]
14. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50:179-211. [doi: 10.1016/0749-5978(91)90020-T]
15. Bandura A. Self-Efficacy: The Exercise of Control. New York: W. H. Freeman; 1997.
16. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
17. Gollwitzer PM. Goal achievement: the role of intentions. Eur Rev Soc Psychol 1993;4(1):141-185. [doi: 10.1080/14792779343000059]
18. Wieber F, Thürmer JL, Gollwitzer PM. Promoting the translation of intentions into action by implementation intentions: behavioral effects and physiological correlates. Front Hum Neurosci 2015 Jul 14;9:395 [FREE Full text] [doi: 10.3389/fnhum.2015.00395] [Medline: 26236214]
19. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
20. Webb TL, Sheeran P. How do implementation intentions promote goal attainment? A test of component processes. J Exp Soc Psychol 2007 Mar;43(2):295-302. [doi: 10.1016/j.jesp.2006.02.001]
21. Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y. An integrated theory of the mind. Psychol Rev 2004 Oct;111(4):1036-1060. [doi: 10.1037/0033-295X.111.4.1036] [Medline: 15482072]
22. Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychol 2010 Jan;29(1):40-49. [doi: 10.1037/a0016993] [Medline: 20063934]
23. Wood W, Neal DT. A new look at habits and the habit-goal interface. Psychol Rev 2007 Oct;114(4):843-863. [doi: 10.1037/0033-295X.114.4.843] [Medline: 17907866]
24. James W. The Principles of Psychology: In Two Volumes. New York: Henry Holt and Company; 1890.
25. Graybiel AM. Habits, rituals, and the evaluative brain. Annu Rev Neurosci 2008;31:359-387. [doi: 10.1146/annurev.neuro.29.051605.112851] [Medline: 18558860]
26. Kahneman D. Thinking, Fast and Slow. New York: Farrar, Straus and Giroux; 2011.

27. Stawarz K, Cox AL, Blandford A. Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. New York, NY, USA: ACM; 2015 Presented at: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul, Republic of Korea p. 2653-2662. [doi: 10.1145/2702123.2702230]
28. Tenison C, Fincham JM, Anderson JR. Phases of learning: how skill acquisition impacts cognitive processing. Cogn Psychol 2016 Jun;87:1-28. [doi: 10.1016/j.cogpsych.2016.03.001] [Medline: 27018936]
29. Sutton RS, Barto AG. Reinforcement Learning. Cambridge, MA: MIT Press; 1998.
30. O'Reilly RC, Frank MJ, Hazy TE, Watz B. PVLV: the primary value and learned value Pavlovian learning algorithm. Behav Neurosci 2007;121(1):31-49. [doi: 10.1037/0735-7044.121.1.31]
31. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement nonreinforcement. In: Black AH, Prokasy WF, editors. Classical Conditioning II: Current Research and Theory. New York: Appleton-Century-Crofts; 1972:64-99.
32. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
33. Riley WT, Martin CA, Rivera DE, Hekler EB, Adams MA, Buman MP, et al. Development of a dynamic computational model of social cognitive theory. Transl Behav Med 2016 Dec;6(4):483-495 [FREE Full text] [doi: 10.1007/s13142-015-0356-6] [Medline: 27848208]
34. Pirolli P. A computational cognitive model of self-efficacy and daily adherence in mHealth. Transl Behav Med 2016 Dec;6(4):496-508 [FREE Full text] [doi: 10.1007/s13142-016-0391-y] [Medline: 27848213]
35. Pirolli P. From good intentions to healthy habits: towards integrated computational models of goals striving and habit formation. Orlando, FL: IEEE; 2016 Presented at: The 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'16); August 16-20, 2016; Orlando, FL p. 181-185. [doi: 10.1109/EMBC.2016.7590670]
36. Pavlik PI, Anderson JR. Using a model to compute the optimal schedule of practice. J Exp Psychol Appl 2008 Jun;14(2):101-117. [doi: 10.1037/1076-898X.14.2.101] [Medline: 18590367]
37. Pavlik Jr PI, Anderson JR. Practice and forgetting effects on vocabulary memory: an activation-based model of the spacing effect. Cogn Sci 2005 Jul 8;29(4):559-586. [Medline: 21702785]
38. Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev 1956 Mar;63(2):81-97. [Medline: 13310704]
39. Simon HA. How big is a chunk?: By combining data from several experiments, a basic human memory unit can be identified and measured. Science 1974 Feb 8;183(4124):482-488. [Medline: 17773029]
40. Tobias R. Changing behavior by memory aids: a social psychological model of prospective memory and habit development tested with dynamic field data. Psychol Rev 2009 Apr;116(2):408-438. [doi: 10.1037/a0015512] [Medline: 19348548]
41. Dismukes RK. Prospective memory in workplace and everyday situations. Curr Dir Psychol Sci 2012;21(4):215-220. [doi: 10.1177/0963721412447621]
42. Elio R. On Modeling Intentions for Prospective Memory Performance. : Cognitive Science Society; 2006 Presented at: 28th Annual Conference of the Cognitive Science Society; July 26-29, 2006; Vancouver, British Columbia, Canada.
43. Taatgen NA. Production compilation: a versatile cognitive mechanism. AISB Quarterly 2004:7.
44. Dayan P. Goal-directed control and its antipodes. Neural Netw 2009 Apr;22(3):213-219. [doi: 10.1016/j.neunet.2009.03.004]
45. Pavel M, Jimison H, Spring B. Behavioral informatics: dynamical models for measuring and assessing behaviors for precision interventions. : IEEE; 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 16-20, 2016; Orlando, FL p. 180-193. [doi: 10.1109/EMBC.2016.7590672]
46. Pirolli PL, Anderson JR. The role of practice in fact retrieval. J Exp Psychol Learn Mem Cogn 1985;11(1):136-153. [doi: 10.1037/0278-7393.11.1.136]
47. Anderson JR, Schooler LJ. Reflections of the environment in memory. Psychol Sci 1991;2(6):396-408. [doi: 10.1111/j.1467-9280.1991.tb00174.x]
48. Gelman A, Carlin J. Beyond power calculations: assessing type S (sign) and type M (magnitude) errors. Perspect Psychol Sci 2014 Nov;9(6):641-651. [doi: 10.1177/1745691614551642] [Medline: 26186114]
49. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
50. 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]
51. Prestwich A, Kellar I. How can the impact of implementation intentions as a behaviour change intervention be improved? Eur Rev Appl Psychol 2014 Jan;64(1):35-41. [doi: 10.1016/j.erap.2010.03.003]
52. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Transl Behav Med 2015 Sep;5(3):335-346. [doi: 10.1007/s13142-015-0324-1]

53. Anderson JR, Boyle CF, Corbett AT, Lewis MW. Cognitive modelling and intelligent tutoring. Artif Intell 1990;42(1):7-49. [doi: 10.1016/0004-3702(90)90093-F]
54. Pirolli P. Cognitive engineering models and cognitive architectures in human-computer interaction. In: Durso FT, Nickerson RS, Schvaneveldt RW, Dumais ST, Lindsay DS, Chi MT, editors. Handbook of Applied Cognition. West Sussex, England: John Wiley & Sons; 1999:441-477.
55. Newell A. Unified Theories of Cognition. Cambridge, MA: Harvard University Press; 1990.
56. Anderson JR. Spanning seven orders of magnitude: a challenge for cognitive modeling. Cogn Sci 2002;26(1):85-112. [doi: 10.1207/s15516709cog2601_3]

Abbreviations

ACT-R: adaptive control of thought-rational
EBI: evidence-based interventions
mHealth: mobile health
OR: odds ratio
SE: standard error
TPB: theory of planned behavior

Edited by G Eysenbach; submitted 21.06.17; peer-reviewed by E Hekler, J Bervoets; comments to author 15.07.17; revised version received 03.09.17; accepted 05.10.17; published 30.11.17.

Please cite as:
Pirolli P, Mohan S, Venkatakrishnan A, Nelson L, Silva M, Springer A
Implementation Intention and Reminder Effects on Behavior Change in a Mobile Health System: A Predictive Cognitive Model
J Med Internet Res 2017;19(11):e397
URL: http://www.jmir.org/2017/11/e397/
doi:10.2196/jmir.8217
PMID:29191800

Original Paper

Health-Seeking Influence Reflected by Online Health-Related Messages Received on Social Media: Cross-Sectional Survey

Rahila Iftikhar[1*], FCPS, MRCGP (Int); Bahaa Abaalkhail[1*], MME, DrPH

Department of Family and Community Medicine, King Abdulaziz University, Jeddah, Saudi Arabia

[*]all authors contributed equally

Corresponding Author:
Rahila Iftikhar, FCPS, MRCGP (Int)
Department of Family and Community Medicine
King Abdulaziz University
College of Medicine
Jeddah, 21589
Saudi Arabia
Phone: 966 6401000 ext 21037
Fax: 966 601141
Email: rahila_iftikhar@hotmail.com

Abstract

Background: Major social networking platforms, such as Facebook, WhatsApp, and Twitter, have become popular means through which people share health-related information, irrespective of whether messages disseminated through these channels are authentic.

Objective: This study aims to describe the demographic characteristics of patients that may demonstrate their attitudes toward medical information shared on social media networks. Second, we address how information found through social media affects the way people deal with their health. Third, we examine whether patients initiate or alter/discontinue their medications based on information derived from social media.

Methods: We conducted a cross-sectional survey between April and June 2015 on patients attending outpatient clinics at King Abdulaziz University, Jeddah, Saudi Arabia. Patients who used social media (Facebook, WhatsApp, and Twitter) were included. We designed a questionnaire with closed-ended and multiple-choice questions to assess the type of social media platforms patients used and whether information received on these platforms influenced their health care decisions. We used chi-square test to establish the relationship between categorical variables.

Results: Of the 442 patients who filled in the questionnaires, 401 used Facebook, WhatsApp, or Twitter. The majority of respondents (89.8%, 397/442) used WhatsApp, followed by Facebook (58.6%, 259/442) and Twitter (42.3%, 187/442). In most cases, respondents received health-related messages from WhatsApp and approximately 42.6% (171/401) reported ever stopping treatment as advised on a social media platform. A significantly higher proportion of patients without heart disease (P=.001) and obese persons (P=.01) checked the authenticity of information received on social media. Social media messages influenced decision making among patients without heart disease (P=.04). Respondents without heart disease (P=.001) and obese persons (P=.01) were more likely to discuss health-related information received on social media channels with a health care professional. A significant proportion of WhatsApp users reported that health-related information received on this platform influenced decisions regarding their family's health care (P=.001). Respondents' decisions regarding family health care were more likely to be influenced when they used two or all three types of platforms (P=.003).

Conclusions: Health education in the digital era needs to be accurate, evidence-based, and regulated. As technologies continue to evolve, we must be equipped to face the challenges it brings with it.

(J Med Internet Res 2017;19(11):e382) doi:10.2196/jmir.5989

KEYWORDS

online health information-seeking behaviors; Facebook; social media; Twitter; WhatsApp

Introduction

Twitter, Facebook, and WhatsApp have become mainstream online tools that permit individuals to connect and share information. Furthermore, they permit individuals to share uncontrolled, unsupervised, and unfiltered content, irrespective of time and place [1]. Consequently, the Internet contains a lot of self-created content [2]. Many people are increasingly using social networking sites for health-related purposes. Research has demonstrated that an increasing number of patients are using social networking sites to share their experiences with health care personnel or institutions [3]. Patients also share their experiences with family members and friends via platforms such as Facebook, WhatsApp, and Twitter [4-6]. There is evidence that if social networking is used properly, then it can help patients [7]. For example, a study that promoted breastfeeding among Saudi women showed increased adherence to breastfeeding through a Twitter campaign [8]. Another study that investigated the effect of Twitter on women's health education demonstrated that women in Saudi Arabia were interested in discussing gynecological complains and breastfeeding-related issues on Twitter [9]. The investigators found that this strategy helped in creating awareness. Moreover, a recent study showed that Twitter was a powerful platform for health promotion strategies [10]. Influential people who have a huge number of followers can constitute an integral part of any health campaign or help in disseminating knowledge.

Concerns about the increasing use of social media to share health experiences and information arise as the use of these sites might affect choices that patients make regarding their health [11]. Furthermore, it might affect the way patients interact with health care professionals. According to one review, information obtained from social networking sites correlated with many measures of quality of care, including performance measures such as mortality and readmission rates [12]. Nevertheless, definitive conclusions cannot be drawn from correlation tests and several questions remain unanswered regarding the impact of patients' use of social media.

There are few data available regarding the impact of social networking sites on the online health information-seeking behaviors of people in Saudi Arabia [13]. This report will attempt to determine whether advice obtained from social media platforms, such as Facebook and WhatsApp, affect choices that people make about their health care. This study aims to describe the demographic characteristics of patients that may demonstrate their attitudes toward medical information shared on social media networks. Second, we address how information found through social media affects the way people deal with their health. Third, we examine whether patients initiate or alter/discontinue their medications based on information derived from social media.

Methods

Participants and Setting

A cross-sectional survey was conducted between April and June 2015 on patients attending King Abdulaziz University Hospital, Jeddah, Saudi Arabia. We included patients who used Facebook, WhatsApp, and Twitter, which are among the most frequently used social media platforms in Saudi Arabia [14]. Informed consent was obtained from all participants prior to recruitment. The Research Ethics Committee at King Abdulaziz University Hospital approved the study.

The initial pool included 442 participants of which 401 reported using Facebook, WhatsApp, or Twitter. The demographics reported reflect the 442 patients who agreed to participate in this survey; however, further analysis was performed only for the 401 participants who used any of the social media platforms under consideration.

Survey Instrument

We modified a previously validated questionnaire [15] to specifically target social media users and how medical information found through social media networks impacts the way they deal with their health. Furthermore, we selected highly prevalent public health issues in Saudi Arabia, such as diabetes mellitus, heart disease, hypertension, and asthma.

The questionnaire was developed with closed-ended and multiple-choice questions that were designed to be nonintrusive and simple to understand. The questionnaire was administered by medical students, who interviewed the participants. Prior to the interview, the students were trained to collect data. Participants were asked to identify their age, gender, nationality, marital status, educational level, and monthly income. They were also asked the types of social media platforms they used and whether information received on these platforms influenced their health care decisions. Returned questionnaires were reviewed and those that were filled in by respondents who did not use any of the three social media platforms (Facebook, WhatsApp, or Twitter) were excluded from analysis.

Statistical Analysis

The data were analyzed using IBM SPSS version 22. Descriptive statistics were computed for all variables. The findings are expressed as counts and percentages for categorical and nominal variables, whereas continuous variables are presented as means and standard deviations. To establish the relationship between categorical variables, chi-square test was used. This test was conducted with the assumption of normal distribution. Lastly, a conventional P value <.05 was adopted to reject the null hypothesis.

Results

Demographic Characteristics

A total of 442 participants with a mean age of 35.4 (11.5) years filled in the questionnaires. Females comprised the majority of the sample (256/442, 71.9%). Approximately 46.0% (193/420) of the respondents had completed at least university education (Table 1). Regarding income, 97 respondents reported incomes greater than 10,000 Saudi riyals (US $2666). Approximately 74.9% (328/438) of the respondents were married and 59.8% (259/433) were Saudis (Table 1).

Type of Social Media Platforms Used

Approximately 90% (397/442) of the respondents used WhatsApp; Twitter was the least used among all three social media platforms. Approximately 32% (142/442) of respondents used all three types of social media platform (Table 2). Respondents received health-related messages more frequently on WhatsApp than Twitter or Facebook. Close to one-third of the respondents reported using all three social media platforms and respondents reported receiving health-related messages more frequently on WhatsApp than on Twitter or Facebook. Despite the number of respondents who reported receiving medical information through social media, less than one-fifth admitted that information shared across these platforms always influenced their health decisions. Further, one-quarter of the respondents admitted to never discussing health-related information with their physicians.

Table 1. Demographic characteristics of the sample (N=442).

Variables	n (%)[a]
Gender	
Male	123 (27.8)
Female	319 (72.2)
Age group (years)	
≤29	138 (31.2)
30-39	123 (27.8)
40-49	104 (23.5)
50-59	49 (11.1)
≥60	28 (6.3)
Marital status	
Married	328 (74.9)
Single	110 (25.1)
Nationality	
Saudi	259 (59.8)
Non-Saudi	174 (40.2)
Educational attainment	
Primary	56 (13.3)
Secondary	158 (37.6)
Graduation/Postgraduation	193 (46.0)
None	13 (3.1)
Occupation	
Housewife	190 (44.8)
Office job	41 (9.7)
Business	19 (4.5)
Doctor	12 (2.8)
Engineer	8 (1.9)
Unemployed	27 (6.4)
Others	127 (30)
Monthly income in Saudi riyals (US$)	
<2000 (533)	46 (11.4)
2000-5000 (533-1333)	135 (33.3)
5000-10,000 (1333-2666)	127 (31.4)
10,000-20,000 (2666-5333)	76 (18.8)
>20,000 (5333)	21 (5.2)

[a]Some cases have missing values.

Table 2. Usage and type of social media used among the respondents (N=442).

Variables	n (%)[a]
Type of social media platform used	
WhatsApp	397 (89.8)
Facebook	259 (58.6)
Twitter	187 (42.3)
Number of social media platforms used	
Never use	41 (9.3)
Any of one	101 (22.9)
Any of two	158 (35.7)
All types	142 (32.1)
Do you receive health-related messages on WhatsApp?	
Yes	311 (78.3)
No	86 (21.7)
Do you receive health-related messages on Facebook?	
Yes	121 (46.7)
No	138 (53.3)
Do you receive health-related messages on Twitter?	
Yes	125 (66.8)
No	54 (28.9)
Not reported	8 (4.3)
Do messages on social media platforms ever influence your decisions regarding you or your family's health care?	
Always	65 (16.2)
Sometimes	185 (46.1)
Never	116 (28.9)
Not reported	35 (8.7)
Do you discuss the authenticity or usefulness of health-related information received on social media platforms with a doctor or other health care professionals?	
Always	128 (31.9)
Sometimes	128 (31.9)
Never	104 (25.9)
Not reported	41 (10.2)

[a]Some cases have missing values, and some respondents used more than one social media platform.

Practices of the Participants Regarding Health-Related Information Received on Social Media Platforms

Among the 401 participants who used Facebook, WhatsApp, or Twitter, less than half admitted starting a treatment as advised on social media without asking their physician (Table 3). In most cases, respondents received health-related messages from WhatsApp and 42.6% (171/401) reported ever stopping treatment as advised on a social media platform. Approximately half (86/171) of respondents were mostly influenced by WhatsApp. Close to one-fifth of the respondents never verified the credibility of the health information received on social media platforms and about one-quarter shared the information without verifying whether it was accurate. Google was cited as the main site where people performed searches to verify the accuracy of health-related information received on social media platforms.

Discussion of Information Received on Social Media With Health Care Professionals Stratified by Health Status

A significantly higher proportion of patients without heart disease (P=.001) and obese persons (P=.01) checked the authenticity of information received through social medial channels (Table 4). There were no differences between persons with diabetes, hypertension, asthma, dyslipidemia, chronic disease, and those without any of these conditions.

Table 3. Practices of social media users regarding information received on social media platforms (N=401).

Variables	n (%)[a]
Have you ever started any medications/treatment as advised/advertised on social media without asking your physician?	
Yes	183 (46.6)
No	210 (53.4)
If yes, which social media platform influenced you most?	
WhatsApp	93 (50.8)
Facebook	24 (13.1)
Twitter	32 (17.5)
Not reported	34 (18.6)
Have you ever stopped any medications/treatment as advised/advertised on social media without asking your physician?	
Yes	171 (42.6)
No	221 (55.1)
Not reported	9 (2.2)
If yes, which social media platform influenced you most?	
WhatsApp	86 (50.3)
Facebook	17 (9.9)
Twitter	29 (17.0)
Not reported	39 (22.8)
Do you verify the credibility of the health information on social media?	
Always	146 (40.7)
Sometimes	149 (41.5)
Never	64 (17.8)
If yes, which sources do you mostly use for verification?	
Google	259 (90.9)
Others (PubMed/Ministry of Health website, etc)	26 (9.1)
Do you verify the credibility of health-related information before sharing it with other people?	
Always	150 (42.3)
Sometimes	116 (32.7)
Never	89 (25.1)

[a]Some cases have missing values.

Table 4. Discussion of information received on social media with health care professionals stratified by health status.

Variables	Discuss the authenticity of information read on social media platforms with health care professionals, n (%)		*P*
	Yes (n=256)	No (n=104)	
Do you have diabetes?			.24
Yes	38 (64.4)	21 (35.6)	
No	199 (72.1)	77 (27.9)	
Do you have heart disease?			.001
Yes	5 (33.3)	10 (66.7)	
No	231 (72.9)	86 (27.1)	
Do you have hypertension?			.46
Yes	35 (76.1)	11 (23.9)	
No	203 (70.7)	84 (29.3)	
Do you have asthma?			.66
Yes	23 (67.6)	11 (32.4)	
No	213 (71.2)	86 (28.8)	
Are you obese?			.01
Yes	42 (87.5)	6 (12.5)	
No	198 (70.0)	85 (30.0)	
Do you have dyslipidemia (increased cholesterol)?			.37
Yes	29 (65.9)	15 (34.1)	
No	211 (72.5)	80 (27.5)	
Do you have any other disease?			.74
Yes	72 (74.2)	25 (25.8)	
No	168 (72.4)	64 (27.6)	
Do you have any chronic disease?			.44
Yes	108 (68.8)	49 (31.2)	
No	140 (72.5)	53 (27.5)	

Association Between Demographic Characteristics and Participants' Attitudes Toward Health-Related Information Shared on Social Media Networks

There was no association between demographics and whether messages ever influenced family health care (Table 5). Married respondents did not differ significantly from single respondents in their attitudes toward medical information shared on social media (*P*=.99). Although participants who had completed secondary education were more likely than primary school leavers and university graduates to be influenced by health messages posted on social media, this difference was not significant (*P*=.52). Similarly, no association was found between respondents' attitudes toward medical information posted on social media and sociodemographic variables, such as occupation (*P*=.95), gender (*P*=.81), nationality (*P*=.53), monthly incomes (*P*=.95), and age (*P*=.31).

Association Between Participants' Health Status and Their Attitudes to Medical Information Shared on Social Media Platforms

Decision making by patients without heart disease was influenced by social media messages (*P*=.04), whereas patients with other chronic diseases were not significantly influenced (*P*=.50). Respondents without heart disease (*P*=.001) and obese persons (*P*=.01) were more likely to discuss health-related information received on social media channels with health care professional. Patients with diabetes, hypertension, asthma, dyslipidemia, or those with any other chronic disease did not differ regarding their likelihood to discuss online health information with health care personnel. Furthermore, persons without diabetes (*P*=.04) or without heart disease (*P*=.001) tended to verify the credibility of information posted on social media channels (Table 6).

Table 5. Association between demographic characteristics and participants' attitudes toward health-related information shared on social media networks[a].

Variables	Messages ever influence decisions regarding family's health care, n (%)			Verify the credibility of the health information on social media from other authentic sources, n (%)		
	Yes (n=250)	No (n=116)	*P*	Yes (n=295)	No (n=64)	*P*
Gender			.08			.58
Male	72 (69.2)	32 (30.8)		82 (80.4)	20 (19.6)	
Female	178 (67.9)	84 (32.1)		213 (82.9)	44 (17.1)	
Age group (years)			.31			.73
≤29	76 (62.3)	46 (37.7)		101 (83.5)	20 (16.5)	
30-39	85 (75.2)	28 (24.8)		88 (80.0)	22 (20.0)	
40-49	59 (67.8)	28 (32.2)		73 (84.9)	13 (15.1)	
50-59	19 (65.5)	10 (34.5)		23 (82.1)	5 (17.9)	
≥60	11 (73.3)	4 (26.7)		10 (71.4)	4 (28.6)	
Marital status			.86			.72
Married	181 (68.3)	84 (31.7)		212 (81.9)	47 (18.1)	
Single	67 (68.4)	31 (31.6)		81 (83.5)	16 (16.5)	
Nationality			.53			.68
Saudi	146 (66.4)	74 (33.6)		176 (81.1)	41 (18.9)	
Non-Saudi	96 (69.6)	42 (30.4)		111 (82.8)	23 (17.2)	
Education			.52			.02
Primary	21 (63.6)	12 (36.4)		20 (64.5)	11 (35.5)	
Secondary	94 (70.1)	40 (29.9)		103 (78.6)	28 (21.4)	
Graduation/Postgraduation	122 (68.9)	55 (31.1)		154 (87.0)	23 (13.0)	
None	1 (33.3)	2 (66.7)		1 (100.0)	0 (0.0)	
Employment			.95			.15
Unemployed	120 (69.8)	52 (30.2)		130 (78.8)	35 (21.2)	
Employed	124 (70.1)	53 (29.9)		151 (84.8)	27 (15.2)	
Monthly income in Saudi riyals (US $)			.95			.36
<5000 (1333)	94 (69.1)	42 (30.9)		109 (80.7)	26 (19.3)	
5000-10,000 (1333-2666)	76 (67.3)	37 (32.7)		93 (86.1)	15 (13.9)	
>10,000 (2666)	61 (68.5)	28 (31.5)		70 (78.7)	19 (21.3)	

[a]The total is <401 in some cases due to missing responses.

Table 6. Association between participants' health status and their attitudes toward information found through social media[a].

Variables	Messages on social media platforms influence decisions regarding family's health care, n (%)			Verify the credibility of the health information on social media from other authentic sources, n (%)		
	Yes (n=250)	No (n=116)	*P*	Yes (n=295)	No (n=64)	*P*
Do you have diabetes?			.66			.04
Yes	38 (65.5)	20 (34.5)		40 (71.4)	16 (28.6)	
No	193 (64.8)	89 (31.6)		230 (83.0)	47 (17.0)	
Do you have heart disease?			.04			.001
Yes	6 (42.9)	8 (57.1)		8 (50.0)	8 (50.0)	
No	222 (68.9)	101 (31.3)		259 (82.5)	55 (17.5)	
Do you have hypertension?			.78			.06
Yes	32 (69.6)	14 (30.4)		31 (70.5)	13 (29.5)	
No	197 (67.5)	95 (32.5)		237 (82.6)	50 (17.4)	
Do you have asthma?			.43			.80
Yes	25 (73.5)	9 (26.5)		29 (82.9)	6 (17.1)	
No	203 (66.8)	101 (33.2)		240 (81.1)	56 (18.9)	
Are you obese?			.30			.68
Yes	34 (73.9)	12 (26.1)		40 (83.3)	8 (16.7)	
No	192 (66.2)	98 (33.8)		227 (80.8)	54 (19.2)	
Do you have dyslipidemia?			.71			.33
Yes	29 (70.7)	12 (29.3)		32 (76.2)	10 (23.8)	
No	203 (67.9)	96 (32.1)		240 (82.5)	51 (17.5)	
Do you have any other disease?			.06			.27
Yes	61 (62.2)	37 (37.8)		79 (79.8)	20 (20.2)	
No	172 (72.6)	65 (27.4)		195 (85.8)	35 (15.2)	
Do you have any chronic disease?			.50			.23
Yes	150 (69.8)	65 (30.2)		121 (79.1)	32 (20.9)	
No	99 (66.4)	50 (33.6)		164 (84.1)	31 (15.9)	

[a]The total is <401 due to missing responses.

Table 7. Association between social media type and participants' attitudes toward medical information shared through social media[a].

Variables	Messages on social media platforms influence decisions regarding family's health care, n (%)			Discuss authenticity of message with health care professionals, n (%)			Verify the credibility of the health information on social media from other authentic sources, n (%)		
	Yes (n=250)	No (n=116)	*P*	Yes (n=256)	No (n=104)	*P*	Yes (n=295)	No (n=64)	*P*
Do you use WhatsApp?			.001			.19			.24
Yes	221 (71.5)	88 (28.5)		221 (72.5)	84 (27.5)		250 (83.1)	51 (16.9)	
No	26 (48.1)	28 (51.9)		33 (63.5)	19 (36.5)		42 (76.4)	13 (23.6)	
Do you use Facebook?			.53			.05			.02
Yes	87 (72.5)	33 (27.5)		91 (75.8)	29 (24.2)		108 (90)	12 (10.0)	
No	84 (68.9)	38 (31.1)		78 (64.5)	43 (35.5)		94 (79)	25 (21.0)	
Do you use Twitter?			.86			.05			.03
Yes	88 (71.0)	36 (29.0)		92 (74.8)	31 (25.2)		109 (89.3)	13 (10.7)	
No	32 (69.6)	14 (30.4)		28 (59.6)	19 (40.4)		36 (76.6)	11 (23.4)	
How many of these social media platforms (Facebook, Twitter, or WhatsApp) do you use?			.003			.12			.10
Only 1	54 (53.5)	47 (46.5)		64 (78)	18 (22.0)		64 (76.2)	20 (23.8)	
2 or all 3	196 (65.3)	104 (34.7)		192 (69.1)	86 (30.9)		231 (84)	44 (16.0)	

[a]The total is <401 due to missing responses.

Impact of Social Media Platform Used on Participants' Attitudes Toward Health-Related Information Shared on These Platforms

A significant proportion of WhatsApp users reported that health-related information disseminated on this platform influenced decisions regarding their family's health care (*P*=.001; Table 7). Similarly, respondents' decisions regarding family health care were more likely to be influenced when they used two or all three types of platforms (*P*=.003). Respondents' decisions regarding family health care did not differ significantly between those who used Facebook or Twitter and those who did not use these platforms.

Discussion

In this study, we explored the impact of health-related information sharing, the influence of social media on peoples' online health information-seeking behavior, and their diligence in following prescriptions, as well as self-medication among social media users. This study shows that most people (89.8%, 397/442) used WhatsApp and 78.3% (311/397) of social media users received health information through these channels. Less than one-fifth of social media users admitted that health-related messages received on these platforms always influenced their decisions regarding family members' health care. Furthermore, a large proportion of patients (46.6%, 186/393) admitted starting medications as advertised on social media platforms without consulting a physician. Similarly, 42.6% (171/401) of patients stopped taking their medication after reading messages received on a social media platform.

It is unquestionable that health care and allied health professionals can use the power of social media to spread information, including recruiting patients for clinical studies and surveying patients to get their opinions on a new treatment or device; however, potential risks may ensue from the use of social media when there are no stringent regulations to share and receive health care information on these platforms. Several investigators have expressed concerns about the potential of social media to negatively impact patients and their treatment [16,17]. In this study, for example, approximately half of respondents who either started or stopped medication were influenced by WhatsApp, reflecting the importance of the way this platform influences how people deal with their health. Less than half the respondents always verified the credibility of information and 90.9% (259/285) performed a Google search to verify the authenticity of messages received through social media channels. Interestingly, 25.1% (89/355) of respondents never discussed health-related messages with their physicians. This might be due to the fact they did not have a regular physician or they did not find it relevant to discuss this with a health care professional. Furthermore, women were more likely to discuss health-related information with their physician for authenticity as compared to men. Another study demonstrated that young male patients sought medical help less frequently and tended to avoid medical consultations [18].

We found that patients who had attained postgraduate college degrees were more likely to verify the credibility of information received via social media channels. According to a previous systematic review [19], educational status appeared to affect the way people evaluated online health information, with individuals with a lower level of education demonstrating worse capacities to evaluate the authenticity of health information shared on social media and lower trust in online health information compared to their more educated peers. Regarding perceived quality of online health information or people's use

of evaluation criteria, the limited number of studies and the diversity of samples and measures do not allow us to draw conclusions about the impact of educational level or other skills-based proxies of health literacy leaving two of the main research questions of this study mainly unanswered. Similarly, we found that patients without diabetes or heart disease were more likely to verify the credibility of medical information shared on social networks. This suggests that patients who have diabetes and heart disease are less likely to verify the authenticity of health information received on social media. This is a concerning factor because patients with chronic diseases should seek medical advice and have regular follow-ups with their doctors and, consequently, should have better education regarding their disease.

The Internet and social media, in particular, provide a business platform to pharmaceutical companies and, according to a recent survey, 40% of top pharmaceutical companies use direct-to-consumer advertising on social media platforms [20]. In the clinical scenario, a physician has to balance the risks against the benefits of prescribing a particular diagnostic test or therapy. Hence, their clinical decisions are based on the patient's understanding, informed consent after explanation of potential risks, preferences, and available resources. On the contrary, social media and other online platforms, which are typically unregulated, may pose a potential threat to patient safety by encouraging the illegal online nonmedical use of prescription drugs [21]. In our setting, for example, we have noticed that patients tend to self-medicate and use complementary medicines. They often get health-related messages on mammography or prostate cancer screening and request to have these investigations without actually understanding the risk or benefit for such diagnostic tests or particular treatment.

Several studies [21-26] have stated that social media has a positive effect on health care, including mental health and physical fitness programs. In fact, it has been suggested that the use of social networking sites to share credible health information, can help physicians fulfill the professional obligation to transmit pertinent information to patients, colleagues, and the public and help members of the public place the findings of health-related current events in proper context. Some physicians affirm that physicians have an ethical obligation to lend their voices to public discourse on health care topics online [27]. Furthermore, it is believed that physicians who use their presence on social media to broadcast their professional commitments and values help fight the unscientific but amplified voices of the media and advertisements, which may disseminate spurious and sometimes dangerously incorrect statements regarding health [28].

We believe that there is an urgent need for mass awareness campaigns to educate people that medical information received on social media channels must be critically reviewed. People should be encouraged to consult their physicians prior to making any self-imposed changes to their prescriptions. Misinformation creates confusion and jeopardizes clinical care. Only 50% of television health shows give evidence-based advice [28], although hosts of television programs perform some degree of research before broadcasting. Similarly, a content analysis of information on urology disseminated on Facebook revealed that only 13% of the posts contained relevant information, whereas 40% were advertisements of commercial products [29]. In the same line, another study that assessed how health conditions were represented on Facebook pages revealed that 32.2% of the information was commercial, whereas 20% were about health awareness [30]. Therefore, it is important that social media users check the authenticity and relevance of all health-related information received on Facebook, Twitter, or WhatsApp. Moreover, there should be cyber surveillance as part of social accountability for spreading potentially incorrect health information. This can be possible by having health professionals edit social network pages to suit patients' needs.

In our context, this study is the first to assess the impact of social media on the way people deal with their health and how messages received on social media platforms influence self-medication practices. However, our findings are limited because this study was conducted at a tertiary care hospital, which may not represent the community setting. Moreover, because we used a cross-sectional convenience sample, we could not establish how social networking affects patients' health decisions.

Our findings indicate that social media is an important tool for health information. In addition, it influences people's behaviors and self-medication practices. This suggests that clinicians need to assess patients' medication histories during every visit. Because compliance to treatment is always an issue for patients with chronic diseases, social media adds another dimension to it. It may provide unauthenticated, misleading information and grounds for unjustified use of medications. Furthermore, the interpretation of messages on social media can be difficult, confusing, and may not be fully comprehended.

Future research should focus on specific diseases such as diabetes mellitus and hypertension, and on the patient's reasons for self-medication. Emphasis should also be placed on the types of medications that patients initiate and those that they stop as well as the consequences associated with such practices. Studies should also explore platforms that patients trust most and how they prefer health information to be communicated to them. Moreover, the reasons underlying people's reluctance to discuss health information and self-medication practices with a physician should be explored in qualitative studies.

Patients should be educated to review all health information skeptically. Policymakers and doctors should endeavor to formulate authenticated local languages, for example, here Arabic health literacy websites where patients can check the credibility of any health-related information received on social media platforms. Health care administrators should also look ahead to plan/forecast future medical care regarding how much and how far doctors wish to be involved in online patient care (digital clinics) [31] and how this will be regularized. Many ethical questions need to be answered before we communicate treatment on Twitter or Facebook.

In conclusion, health education in the digital era needs to be accurate, evidence-based, and regulated. As technology continues to evolve, we must be equipped to face the challenges it brings with it. The two main challenges in this regard include

legislation and patient confidentiality. Social media cannot replace proper consultation, listening to nonverbal cues, touch, physical examination, exploring patients' ideas, expectations, and individualized care. Therefore, decisions regarding major clinical care should be encouraged in the professional setting.

Acknowledgments

The authors would like to thank the medical students Dr Zainab Abdulelah Alrashed, Dr Basail Naji Ali Alnahab, Dr Hawraa Ismaaiel Mosa Alhmied, Dr Fatimah Salman Alnazghan, Dr Arwa Ahmed Alhajji, Dr Alawayah Jawad Alahmed, Dr Wall Hussain Aljumeal, and Dr Zahra Malik Alsadiq for their efforts in helping with data collection.

Conflicts of Interest

None declared.

References

1. Hoffman DL, Novak TP. Why do people use social media? Empirical findings and a new theoretical framework for social media goal pursuit. SSRN 2012:33 [FREE Full text] [doi: 10.2139/ssrn.1989586]
2. Van De Belt TH, Engelen LJLPG, Berben SAA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
3. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
4. Adams S. Post-panoptic surveillance through healthcare rating sites. Inform Commun Soc 2013 Mar;16(2):215-235 [FREE Full text] [doi: 10.1080/1369118X.2012.701657]
5. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med 2011 Mar;26(3):287-292 [FREE Full text] [doi: 10.1007/s11606-010-1526-3] [Medline: 20945113]
6. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
7. 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]
8. Bahkali S, Alkharjy N, Alowairdy M, Househ M, Da'ar O, Alsurimi K. A social media campaign to promote breastfeeding among Saudi women: a web-based survey study. Stud Health Technol Inform 2015;213:247-250. [Medline: 26153006]
9. Bahkali S, Almaiman A, Bahkali A, Almaiman S, Househ M, Alsurimi K. The role of social media in promoting women's health education in Saudi Arabia. Stud Health Technol Inform 2015;213:259-262. [Medline: 26153009]
10. Albalawi Y, Sixsmith J. Identifying Twitter influencer profiles for health promotion in Saudi Arabia. Health Promot Int 2015 Oct 29:pii. [doi: 10.1093/heapro/dav103] [Medline: 26516181]
11. Greaves F, Pape U, King D. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012;21(7):5. [doi: 10.1136/bmjqs-2012-000906]
12. Verhoef LM, van de Belt TH, Engelen LJ, 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]
13. Almaiman S, Bahkali S, Al Farhan A, Bamuhair S, Househ M, Alsurimi K. The prevalence of using social media among healthcare professionals in Saudi Arabia: a pilot study. Stud Health Technol Inform 2015;213:263-266. [Medline: 26153010]
14. Statista. 2017. Penetration of leading social networks in Saudi Arabia as of 4th quarter 2015 URL: https://www.statista.com/statistics/284451/saudi-arabia-social-network-penetration/ [accessed 2017-10-05] [WebCite Cache ID 6tzY2e7mw]
15. Van de Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013 Oct;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
16. Diviani N, van den Putte B, Giani S, van Weert J. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018]
17. Coulson N. How do online patient support communities affect the experience of inflammatory bowel disease? An online survey. JRSM Short Rep 2013;4(8):2042533313478004 [FREE Full text] [doi: 10.1177/2042533313478004] [Medline: 24040493]
18. Smailhodzic E, Hooijsma W, Boonstra A, Langley D. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]

19. Zarcadoolas C, Pleasant A, Greer D. Advancing Health Literacy: A Framework for Understanding and Action. San Francisco, CA: Jossey-Bass; 2006.
20. Tyrawski J, DeAndrea D. Pharmaceutical companies and their drugs on social media: a content analysis of drug information on popular social media sites. J Med Internet Res 2015 Jun 01;17(6):e130 [FREE Full text] [doi: 10.2196/jmir.4357] [Medline: 26032738]
21. Mackey TK, Liang BA, Strathdee SA. Digital social media, youth, and nonmedical use of prescription drugs: the need for reform. J Med Internet Res 2013 Jul 26;15(7):e143 [FREE Full text] [doi: 10.2196/jmir.2464] [Medline: 23892156]
22. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
23. Shepherd A, Sanders C, Doyle M, Shaw J. Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation. BMC Psychiatry 2015;15:29 [FREE Full text] [doi: 10.1186/s12888-015-0408-y] [Medline: 25881089]
24. Williams JP, Schroeder D. Popular glucose tracking apps and use of mHealth by Latinos with diabetes: review. JMIR Mhealth Uhealth 2015;3(3):e84 [FREE Full text] [doi: 10.2196/mhealth.3986] [Medline: 26307533]
25. Kear T, Harrington M, Bhattacharya A. Partnering with patients using social media to develop a hypertension management instrument. J Am Soc Hypertens 2015 Sep;9(9):725-734 [FREE Full text] [doi: 10.1016/j.jash.2015.07.006] [Medline: 26369442]
26. Koster ES, Philbert D, de Vries TW, van Dijk L, Bouvy ML. "I just forget to take it": asthma self-management needs and preferences in adolescents. J Asthma 2015 Oct;52(8):831-837. [doi: 10.3109/02770903.2015.1020388] [Medline: 26037397]
27. Hoeksma J. ehealth Insider. Call to join patients online URL: https://www.ehilive.co.uk/en/home.html [accessed 2017-10-05] [WebCite Cache ID 6tzZcnsmo]
28. Korownyk C, Kolber MR, McCormack J, Lam V, Overbo K, Cotton C, et al. Televised medical talk shows--what they recommend and the evidence to support their recommendations: a prospective observational study. BMJ 2014 Dec 17;349:g7346 [FREE Full text] [Medline: 25520234]
29. Sajadi KP, Goldman HB. Social networks lack useful content for incontinence. Urology 2011 Oct;78(4):764-767. [doi: 10.1016/j.urology.2011.04.074] [Medline: 21862115]
30. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. J Med Internet Res 2014;16(8):e182 [FREE Full text] [doi: 10.2196/jmir.3275] [Medline: 25092386]
31. Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AY, et al. Ethical issues of social media usage in healthcare. Yearb Med Inform 2015 Aug 13;10(1):137-147 [FREE Full text] [doi: 10.15265/IY-2015-001] [Medline: 26293861]

Edited by G Eysenbach; submitted 18.05.16; peer-reviewed by Q Yang, A Pathipati, H Johnson; comments to author 31.07.16; revised version received 21.10.16; accepted 15.08.17; published 16.11.17.

Please cite as:
Iftikhar R, Abaalkhail B
Health-Seeking Influence Reflected by Online Health-Related Messages Received on Social Media: Cross-Sectional Survey
J Med Internet Res 2017;19(11):e382
URL: http://www.jmir.org/2017/11/e382/
doi:10.2196/jmir.5989
PMID:29146568

Original Paper

Using Facebook to Recruit Young Australian Men Into a Cross-Sectional Human Papillomavirus Study

Roopa Das[1,2], BBiomed, MD; Dorothy A Machalek[2,3], MPH, PhD; Edmund G Molesworth[1,2], MD; Suzanne M Garland[2,3,4,5], MBBS, MD, FRCPA, FRANZCOG Ad Eundem, FRANZCOG, FAChSHM

[1]Melbourne Medical School, University of Melbourne, Parkville, Australia

[2]Murdoch Children's Research Institute, The Royal Children's Hospital, Parkville, Australia

[3]Department of Microbiology and Infectious Diseases, The Royal Women's Hospital, Melbourne, Australia

[4]Department of Obstetrics and Gynaecology, University of Melbourne, Parkville, Australia

[5]The Royal Children's Hospital, Parkville, Australia

Corresponding Author:
Suzanne M Garland, MBBS, MD, FRCPA, FRANZCOG Ad Eundem, FRANZCOG, FAChSHM
Department of Microbiology and Infectious Diseases
The Royal Women's Hospital
20 Flemington Rd
Melbourne, 3052
Australia
Phone: 61 383453670
Fax: 61 393478235
Email: suzanne.garland@thewomens.org.au

Abstract

Background: Young men can be difficult to engage in health research using traditional methods of recruitment. Social networking sites are increasingly being used to recruit participants into health research, due to their cost effectiveness, overall generalizability, and wide reach.

Objective: The aim of this study was to determine the feasibility of using Facebook to recruit young Australian men into a human papillomavirus (HPV) prevalence study.

Methods: We recruited male permanent residents of Australia, aged 18 to 35 years, into the HPV in Young Males (HYM) study through targeted advertising placed on Facebook. Consenting participants completed an online questionnaire and provided a self-collected penile swab for HPV DNA detection and genotyping. We compared sociodemographic characteristics of the study population with those of the general Australian male population, based on Australian 2011 census data.

Results: Between February 2015 and February 2017, targeted Facebook advertisements reached 1,523,239 men, resulting in 41,811 clicks through to the study website, with 1072 (2.56%) converting to lodgment of an expression of interest. Of these, 681 (63.53%) provided written informed consent and 535 (78.6% of recruited participants) completed all the study requirements. Reasons for participating in the study included altruism, past history of HPV, gaining more knowledge about HPV or the vaccine, working in the health industry, and the monetary compensation. The average advertising cost per completed study participant was Aus $48. Compared with the census population, HYM study participants were more likely to be Australian born (P<.001), be from Victoria (P=.003) or the Australian Capital Territory (P=.004), reside in a major city (P<.001), and have completed undergraduate (P<.001) or postgraduate education (P<.001). HYM study participants were less likely to report being a current smoker (P=.03), but were more likely to identify as bisexual or homosexual (294/529, 55.6%, P<.001), than the general population.

Conclusions: Using Facebook is a feasible and efficient strategy for the recruitment of men from across Australia for HPV testing. This method could be used for monitoring the impact of HPV vaccination. Additional targeting may achieve a sample that is broadly demographically representative of the Australian population. Future research should explore how the sexual risk behavior characteristics of populations recruited through Facebook compare with those of traditional recruitment methods.

(J Med Internet Res 2017;19(11):e389) doi:10.2196/jmir.8739

KEYWORDS

social media; Facebook; human papillomavirus; HPV; online recruitment; social networking; social networking sites; SNS

Introduction

Human papillomavirus (HPV) is one of the most common sexually transmitted infections globally. Infection with HPV is mostly asymptomatic; however, its clinical sequelae include genital warts, cervical cancer, and less commonly, a proportion of vaginal, vulvar, oral, oropharyngeal, anal, and penile cancers [1-3]. Many countries have introduced female HPV vaccination into their national immunization programs, with strong evidence showing reductions in vaccine-related-type HPV infections and disease where vaccine uptake has been high [4-6]. However, despite demonstrated trial efficacy [7,8] and the growing evidence that HPV also causes cancers in men (ie, penile, anal, and oropharyngeal), there is ongoing international debate on whether boys should be included in routine vaccination programs [9-14]. Consequently, only a handful of countries, including Australia, have implemented sex-neutral HPV vaccination approaches [15,16]. Monitoring the impact of HPV vaccination in men in these settings is thus important to document reductions in HPV-related biological end points (ie, genital HPV infections and cancers) caused by vaccine-targeted HPV genotypes and to inform whether routine vaccination of boys will result in any benefits above those that have already been achieved through existing female vaccination programs [17,18].

Measuring rates of genital HPV infection in men is challenging. HPV is very common, typically asymptomatic, and usually clears spontaneously. Furthermore, cancers associated with these infections are very rare. Thus, there are no population-based registries or screening programs, which could form the sampling frame for HPV vaccine monitoring studies in the general male population [19]. In view of this, to date, most direct estimates of HPV prevalence have been achieved through dedicated studies, whereby specimens for HPV testing are collected from opportunistic sampling of males recruited from clinical or community-based settings [17,20,21]. However, such studies are often expensive and require significant resources. Linking in with existing national surveillance activities, such as those implemented as part of the National Health and Nutrition Examination Surveys in the United States [22], is an ideal option for easily reaching populations of interest. However, in the absence of such frameworks, alternative and sustainable strategies are needed to collect genital specimens for the purpose of monitoring the impact of HPV vaccination on men in the long term.

Social networking sites (SNSs) present new opportunities for recruiting young men for health research, and they may have greater reach, be less time consuming, and be cheaper than traditional recruitment methods [23,24]. An estimated 99% of 18- to 29-year-old Australians use the social networking site Facebook, with 75% accessing this site at least once a day [25]. Facebook is also accessed throughout metropolitan and regional Australia, with at least 92% of people from each state of Australia using the platform [25]. This highlights the potential for access to a demographically and geographically diverse sample of participants. Advertising through SNSs has been shown to be successful for the recruitment of young females for surveys including health promotion and sexual health screening [24,26-31]. While SNSs have been effective in recruiting males into trials and sexual health promotion campaigns and surveys, both in Australia and internationally, including a recent US-based feasibility study aimed at recruiting males into a randomized controlled trial of HPV vaccination [32], little has been published with respect to study designs involving the collection of clinical specimens [33,34].

Our group has previously shown that Facebook was a cost-effective and fast method of recruiting young Australian women for home-based HPV testing [26]. In this study, we assessed the feasibility of using targeted Facebook advertising to recruit healthy Australian men aged 18 to 35 years into a HPV prevalence study.

Methods

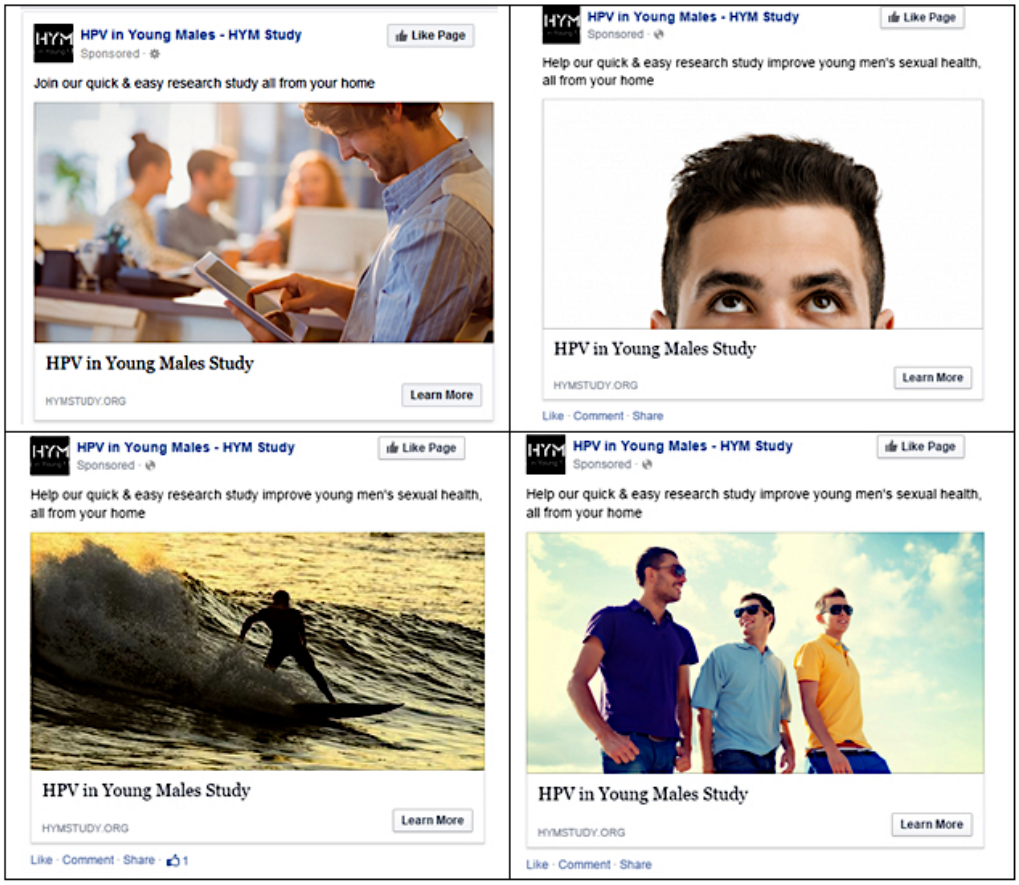
HPV in Young Males (HYM) is an ongoing study of HPV genotype prevalence in Australian men aged 18 to 35 years, recruited via Facebook. The study is part of a larger national surveillance program aiming to monitor circulating HPV genotypes in the Australian population, to evaluate the effectiveness of the Australian National HPV Vaccination Program [17]. The HYM study protocol was approved by the Royal Women's Hospital Human Research and Ethics Committee (HREC Project number 14/22).

Recruitment

Recruitment to the study was based on strategies developed for the Vaccine Against Cervical Cancer Impact and Effectiveness (VACCINE) study, as previously described [6,35]. We recruited participants through targeted advertisements placed on Facebook (Facebook, Inc) between February 2015 and February 2017. Advertisements were targeted to men who were aged 18 to 35 years and living in Australia (we specifically selected the audience criteria on the Ads Manager of Facebook: users with an Australian IP address, male, and aged 18-35 years). Each advertisement contained a short text and image, examples of which are presented in Figure 1. Once clicking on the advertisement, participants were directed to the HYM password-protected website. Here potential participants could read about the study and register their expression of interest (EOI). Only men who registered their EOI were contacted by a study researcher. A telephone call established eligibility and obtained verbal consent. The inclusion criteria confirmed verbally were (1) male, (2) aged 18 to 35 years, (3) living permanently in Australia and eligible for a Medicare number, and (4) had been sexually active (ie, had ever had penetrative or oral sex).

After we obtained verbal consent, we sent a link to an online consent form to participants via SurveyMonkey (SurveyMonkey Inc). The consent consisted of 2 parts: consent for the study and consent to confirm the participant's vaccination status with the Australian National HPV Vaccination Program Register. Once the online consent was completed, an online questionnaire and swab pack was sent out to participants. The online questionnaire, also done through SurveyMonkey, covered questions about basic demographics, sexual history, and knowledge and attitudes toward HPV and the HPV vaccines.

Figure 1. Examples of Facebook recruitment advertisements for the HPV in Young Males (HYM) study, as seen on News Feeds. HPV: human papillomavirus.



The swab pack contained everything needed to perform a self-collected penile swab in the privacy of a participant's home. Participants were able to request and obtain their HPV DNA result. Once participants completed all the study requirements, they were compensated with an Aus $20 gift voucher for their time and effort. Participants were made aware of this compensation via the study website and during verbal consent. Among participants with outstanding study requirements, 3 reminders were sent via email, before the person was deemed lost from the study.

Statistical Analysis

We used Facebook metrics to calculate the amount spent on the advertising campaign. Facebook used a bidding system to display advertisements, with a cost allocated each time an advertisement was clicked [24]. Demographics for analyses were as follows: age (grouped as 18-24 years, 25-29 years, and 30-35 years), Australian state or territory of residence, gender identity, education, geographical remoteness (major city, inner regional, or outer regional or remote [36]), Aboriginal or Torres Strait Island status, country of birth, relationship status (single or casual, or committed relationship, which included marriage and de facto relationships), smoking, income, and employment status. We compared demographics with the 2011 census data on men aged 18 to 35 years, from the Australian Bureau of Statistics [37], and the Second Australian Study of Health and Relationships published by the Commonwealth Scientific and Industrial Research Organisation [38]. Chi-square, 95% CIs, and P values <.05 were used to assess significance. Stata 14 (StataCorp LLC) was used to perform statistical analyses.

Results

Between February 2015 and February 2017, we simultaneously ran one campaign containing 5 advertisements, which reached 1,523,239 Australian male Facebook users within the target age range. In that time, the campaign was presented to users 7,029,581 times. The advertisements were seen on mobile phones or tablets (n=687,682, 45.15%), desktop computers (n=336,813, 22.11%), and other devices (n=498,744, 32.74%). There were 41,811 clicks (0.59% of total advertisement presentations) on the advertisements, resulting in 1072 (2.56% of clicks) EOIs. The advertisements were shown at all times and were viewed on average 4.61 times per day (the number varied depending on the amount of money placed on the advertisement).

Of the 1072 participants who submitted an EOI, 681 (63.53% of EOIs) provided verbal and written consent and were recruited into the study. The primary reasons for expressing an interest to participate in the study were altruism (n=577, 53.82%); being interested in the area (ie, sexual health, HPV or vaccination) (n=180, 16.79%); working in medical research or a health-related area (n=129, 12.03%); wanting to know their HPV result (n=64, 5.97%); and having a history of, or knowing someone with a history of, HPV infection or related disease (eg, genital warts or cervical lesions) (n=96, 8.96%). Only 2.42%

(n=26) of men reported the monetary compensation as a reason for wanting to participate.

Of the remaining 391 (36.47%) EOIs, the majority (n=352, 90.0%) could not be contacted to obtain verbal or written consent and were subsequently lost from the study. A further 37 (9.5%) were deemed ineligible after telephone contact, and 2 (0.5%) were yet to be contacted. The total amount of money spent on the advertisement campaign in this time period was Aus $25,687.64, with the average cost per click being Aus $0.61. The cost per EOI translated to Aus $23.96. The cost per completed participant was Aus $48.01.

Table 1 presents the demographic characteristics of the 1072 participants who submitted an EOI. The median age of respondents was 26 years (interquartile range 23-30 years). The vast majority of EOIs (n=929, 86.66%) were generated directly from clicking on Facebook advertisements. Of the remaining 143 EOIs, 84 (7.8% of the total) were derived from referrals by friends or family, 20 (1.9%) from media (eg, Flipboard), and 39 (3.6%) from other sources. The vast majority of men who submitted an EOI resided in the eastern states of Australia (882/1059, 83.29%). There were no significant differences in age, how participants heard of the study, and the state in which they resided between respondents who were recruited into the study and those who were not.

Of the 681 recruited participants, 535 (78.6%) had completed the study at the time of writing this report. The median time from the date of EOI to the date that the swab pack was returned was 5 weeks (interquartile range 3-8). Of the remaining 146 participants, 91 (13.3%) were in the process of completing the study, and 55 (8.1%) did not return the swab pack after multiple reminders and were therefore deemed lost to follow-up. There were no differences in age between participants who completed the study and those who did not return their swab pack.

Table 2 presents the sociodemographic characteristics of the 535 HYM study participants who completed the study compared with those of the general Australian age-matched male population. Compared with the general population, there was a significantly higher proportion of men aged 25 to 29 years of age participating in the study (*P*=.003). There was also a significant difference in location of residence between men who completed the study and the general population. Men residing in Victoria (*P*=.003) and the Australian Capital Territory (*P*=.004) were overrepresented in our cohort, while those living in Western Australia (*P*<.001) were underrepresented.

Table 1. Demographic characteristics of 1072 men who expressed interest to participate in the HPV[a] in Young Males (HYM) study.

Characteristics	Overall (N=1072)[b]		Recruited into HYM[c]		Not recruited into HYM (n=391)		*P* value[d]
	n	%	n	%	n	%	
Age group (years)							.07
18-24	384	36.23	240	35.2	144	38.0	
25-29	376	35.47	232	34.1	144	38.0	
30-35	300	28.30	209	30.7	91	24.0	
Heard about the study from...							.09
Facebook ad	929	86.66	590	86.6	339	86.7	
Friend or family	84	7.84	59	8.7	25	6.4	
Media	20	1.86	8	1.2	12	3.1	
Other[e]	39	3.64	24	3.5	15	3.8	
Australian state of residence							.56
Victoria	318	30.03	201	29.5	117	31.0	
New South Wales	311	29.37	203	29.8	108	28.6	
Queensland	216	20.40	143	21.0	73	19.3	
Western Australia	85	8.03	47	6.9	38	10.0	
South Australia	59	5.57	37	5.4	22	5.8	
Australian Capital Territory	37	3.49	25	3.7	12	3.2	
Tasmania	22	2.08	17	2.5	5	1.3	
Northern Territory	11	1.03	8	1.2	3	0.8	

[a]HPV: human papillomavirus.

[b]Numbers do not always add up to 1072 due to small amount of missing data.

[c]Participants who provided verbal and written consent.

[d]Chi-square test was used to determine *P* value.

[e]Other includes hearing about the study from university, work, or health care settings.

Table 2. Demographic characteristics of 535 HPV[a] in Young Males (HYM) study participants compared with the general Australian male population aged 18-35 years.

Characteristics	Completed the study (n=535)[b]			General population (%)[c]	*P* value[d]
	n	%	95% CI		
Age group (years)					
18-24 years	193	36.1	32.0-40.3	38.9	.18
25-29 years	184	34.4	30.4-38.6	28.5	.003
30-35 years	158	29.5	25.7-33.6	32.6	.13
Australian state of residence					
Victoria	167	31.2	27.3-35.3	25.5	.003
New South Wales	161	30.1	26.2-34.2	31.3	.55
Queensland	109	20.4	17.0-24.0	19.9	.77
Western Australia	33	6.2	4.3-8.6	11.1	<.001
South Australia	24	4.5	2.9-6.6	7.1	.02
Australian Capital Territory	20	3.7	2.3-5.7	2.0	.004
Tasmania	14	2.6	1.4-4.4	1.9	.24
Northern Territory	7	1.3	0.5-2.7	1.2	.86
Education					
Year 12 or below	110	20.6	17.2-24.2	45.9	<.001
Undergraduate or certificate	340	63.6	59.3-67.6	50.1	<.001
Postgraduate	79	14.8	11.9-18.1	4.0	<.001
Geographic region					
Major cities	441	82.4	78.9-85.6	75.1	<.001
Inner regional	54	10.1	7.7-13.0	14.8	.002
Outer regional or remote	40	7.5	5.4-10.0	10.1	.045
Aboriginal or Torres Strait Islander					
Yes	15	2.8	1.6-4.6	2.9	.92
No	514	97.2	95.3-98.4	97.1	.92
Country of birth					
Australia	439	83.0	79.5-86.1	72.8	<.001
Other	90	17.0	13.9-20.5	27.2	<.001
Employment status					
Full-time	306	57.2	53.7-62.3	61.7	.09
Part-time	141	26.4	23.0-30.8	16.6	<.001
Not in the labor force[e]	47	8.8	6.6-11.7	14.9	<.001
Unemployed	33	6.2	4.3-8.7	6.8	.62
Gender identity					
Heterosexual	235	44.4	40.1-48.8	96.7	<.001
Bisexual or homosexual	294	55.6	51.2-59.9	3.3	<.001
Relationship status					
Single or casual	256	49.2	44.5-53.2	60.6	<.001
Committed	272	50.8	471-55.9	39.4	<.001
Regular cigarette smoker					
Never	353	66.7	62.5-70.7	53.7	<.001

Characteristics	Completed the study (n=535)[b]			General population (%)[c]	*P* value[d]
	n	%	95% CI		
Past	94	17.8	14.6-21.3	26.4	<.001
Current	82	15.5	12.5-18.9	19.9	.03

[a]HPV: human papillomavirus.

[b]Numbers do not always add up to 535 due to small amount of missing data.

[c]Target population data sourced from the 2011 census by the Australian Bureau of Statistics and Second Australian Study of Health and Relationships [36,37].

[d]Chi-square test was used to determine *P* values.

[e]Not in the labor force and not looking for work included full-time caregivers or students.

A larger proportion of HYM study participants had higher education (*P*<.001), lived in major cities in Australia (*P*<.001), were born in Australia (*P*<.001), had never smoked cigarettes or had quit smoking (*P*<.001 never and past; *P*=.03 current), were in a committed relationship (*P*<.001), or worked part-time (*P*<0.001). Study participants were also more likely to identify as gay or bisexual than the underlying population (*P*<.001). There was no significant difference in indigenous status between the 2 populations (*P*=.92).

Discussion

Principal Findings

In this study, we used targeted Facebook advertising to recruit Australian men aged 18 to 35 years into a cross-sectional HPV genotype prevalence study. Consenting participants were required to complete an online questionnaire and return a self-collected penile swab for HPV detection and genotyping. Over a 2-year study period, 64% of the 1072 men who submitted an EOI were recruited, and of these, the majority (79%) completed the study requirements, making this an acceptable and effective strategy for collecting specimens, which are not routinely collected clinically. However, while the strategy was convenient and had the potential to reach a large cross-section of the population, the sociodemographic characteristics of men who participated in the HYM study were proportionally different from those of the general population of the same age as measured by the census, suggesting that additional targeting is needed to achieve a sample that is broadly demographically representative of the Australian population.

Online recruitment methods have been used for behavioral surveys among men, but few studies have reported their use for subsequent self-collection of a specimen [33,34,39]. The HYM study findings suggest that Facebook can be used to recruit young men for HPV testing and that social media could provide a potential alternative avenue for targeting young men for sexually transmitted infection screening in general. A recent Australian study showed that home-based sexually transmitted infection screening was popular among young men due to such screening being free and convenient [39]. Our study had a high return rate of self-collected specimens (79%), highlighting the feasibility of sending swabs for self-collection by mail. The high return rates are in contrast to those reported in some published studies using home-based collection kits for sexually transmitted infection screening (48.2% in Australia, 40% in Sweden, 7.8% in the United States, and 33% in the Netherlands) [39-42]. This may reflect differences in study design, namely, the option of returning specimens by mail, rather than directly to a collection center, in addition to providing multiple reminders to return the pack, rather than none or only 1 reminder.

While using Facebook was an efficient strategy for targeting men for recruitment, the cost per participant recruited was Aus $24 higher than with women participating in a similar HPV vaccine effectiveness study [26]. Our study reached more men than the number of women reached by the VACCINE study (1,523,239 men and 984,159 women), despite the shorter duration of our campaign, and had a higher click-through rate to the study website (0.6% compared with 0.04% for the VACCINE study's female participants). This likely reflected the broader eligibility criteria of this study, with all Australian men aged 18 to 35 years targeted for recruitment, compared with the VACCINE study, where recruitment was limited to 18- to 25-year-old women living in Victoria [26]. Despite this, once on the study website, a lower proportion of men submitted an EOI compared with the female study (3% versus 6% respectively). We did not collect reasons for not wanting to participate in the HYM study; however, we hypothesize that the potentially confronting study requirements may have, to some extent, discouraged some men from taking part. Self-collecting a sample of cells from the surface of the penis is an acceptable and common form of specimen collection for HPV testing in epidemiological studies [20,43-45]. However, the method is not routine clinical practice and is likely to be unfamiliar to most men [46].

While to the best of our knowledge HYM is the first study to recruit young Australian men into an HPV prevalence survey through advertising on SNSs, elements of the study findings, with respect to cohort characteristics, are consistent with previous published research. For example, we found that our cohort was more educated than the general population. Studies recruiting through SNSs commonly recruit cohorts that are more highly educated than the general population [24,26,28,47]. Likewise, our cohort consisted of a larger proportion of younger men, which is consistent with the population of users of SNSs, who tend to be younger than the general population [25]. However, more men from Victoria and the Australian Capital Territory, and fewer men from Western Australia, submitted an EOI and completed the study. Victorian men may have been more inclined to participate, as this study was based in Victoria.

Further, logistical reasons such as the time difference between the eastern states and Western Australia posed a large difficulty in receiving verbal consent, contributing to lower numbers of participants from Western Australia.

Similarly to previous research, we found that a significantly higher proportion of men who identified as homosexual or bisexual participated in the study [32], despite the broad eligibility criteria and attempts to achieve sexual preference neutrality in both the images and text displayed on the recruitment advertisements (see Figure 1). The use of Internet-based social media has been widely promoted as a feasible strategy for the recruitment of men who have sex with men. Their overrepresentation observed in the HYM study may have been due to heightened knowledge and concern regarding HPV among this population, who are at increased risk of HPV-related disease [42,43,48]. This highlights the need for additional targeted advertising to achieve a sample that is broadly demographically representative of the general population. A benefit of using Facebook is the ability to readily modify how one targets specific cohorts for recruitment by adjusting the wording, images, and placement of the advertising, as has been shown in previous studies [26,49].

The main question arising from the study, with respect to the suitability of using SNSs such as Facebook for long-term HPV monitoring, is the potential impact of this form of convenience sampling on the stability of the population under surveillance, as well as the external validity of any observed HPV prevalence estimates. Future research should further explore how the demographic and sexual risk behavior characteristics of populations recruited through Facebook compare with traditional recruitment methods (ie, clinical or community-based samples), and monitor how samples of young men recruited via Facebook change over time with respect to demographic and sexual behavior characteristics.

Furthermore, it is important to note that, while Facebook is the most popular SNS in Australia, the use of other SNSs such as Twitter, Instagram, and Snapchat is on the rise [25]. Of note, SNSs that use imaging and visual components such as Snapchat and Instagram are particularly popular among younger Australians aged less than 30 years. Therefore, it will be interesting in the future for studies to assess the effectiveness of recruiting younger participants through other SNSs.

Conclusions

This study demonstrated that using Facebook was an effective strategy in recruiting young Australian men for home-based HPV testing, with high return rates of specimens. This modality could potentially be used to monitor the prevalence of circulating HPV genotypes in Australia to help determine the impact of routine HPV vaccination. Moreover, home-based swab collection is a potential avenue for future health screening among men.

Our study population was different from the general population. However, through targeting of Facebook advertisements, and collecting demographic and sexual behavior data, we can hope to mitigate this issue in future analyses. Further research is needed in assessing the demographic characteristics of men recruited through Facebook over time and compared with men recruited through more traditional methods to assess the viability of using Facebook in long-term HPV monitoring. With the exponentially fast rate of change among SNS uptake, more research is needed to inform the feasibility of using other SNSs to recruit men into health studies.

Acknowledgments

The HPV in Young Males study is funded by a National Health and Medical Research Council Program Grant (Sexually transmitted infections: Causes, consequences and interventions Grant #568971) and an Australian Government (Department of Health) HPV Surveillance Fund (Grant #H1314G010). We would like to thank all the staff who contributed to recruitment of participants and data collection, laboratory staff, and all the participants for their time and effort.

Conflicts of Interest

SMG reports grants from Merck Sharp & Dohme (MSD), GlaxoSmithKline plc, CSL Limited, and the Commonwealth Department of Health outside the submitted work; has received nonfinancial support from Merck; and has delivered lectures and received speaking fees from Merck Sharp & Dohme and Sanofi Pasteur MSD for work performed in her personal time. DAM has received honoraria and travel funding from Merck and travel funding from Seqirus (formerly bioCSL), outside the submitted work.

References

1. Koutsky L. Epidemiology of genital human papillomavirus infection. Am J Med 1997 May 05;102(5A):3-8. [Medline: 9217656]
2. Georgousakis M, Jayasinghe S, Brotherton J, Gilroy N, Chiu C, Macartney K. Population-wide vaccination against human papillomavirus in adolescent boys: Australia as a case study. Lancet Infect Dis 2012 Aug;12(8):627-634. [doi: 10.1016/S1473-3099(12)70031-2] [Medline: 22445354]
3. Centers for Disease Control an Prevention. Prevention of genital HPV infection and sequelae: report of an external consultants' meeting. 1999 Dec. URL: https://www.cdc.gov/std/hpv/hpvsupplement99.pdf [accessed 2017-07-07] [WebCite Cache ID 6rmBN9eQy]

4. Tabrizi SN, Brotherton JML, Kaldor JM, Skinner SR, Liu B, Bateson D, et al. Assessment of herd immunity and cross-protection after a human papillomavirus vaccination programme in Australia: a repeat cross-sectional study. Lancet Infect Dis 2014 Oct;14(10):958-966. [doi: 10.1016/S1473-3099(14)70841-2] [Medline: 25107680]
5. Tabrizi SN, Brotherton JML, Kaldor JM, Skinner SR, Cummins E, Liu B, et al. Fall in human papillomavirus prevalence following a national vaccination program. J Infect Dis 2012 Dec 01;206(11):1645-1651. [doi: 10.1093/infdis/jis590] [Medline: 23087430]
6. Young EJ, Tabrizi SN, Brotherton JM, Wark JD, Pyman J, Saville M, et al. Measuring effectiveness of the cervical cancer vaccine in an Australian setting (the VACCINE study). BMC Cancer 2013;13:296 [FREE Full text] [doi: 10.1186/1471-2407-13-296] [Medline: 23777549]
7. Hillman RJ, Giuliano AR, Palefsky JM, Goldstone S, Moreira ED, Vardas E, et al. Immunogenicity of the quadrivalent human papillomavirus (type 6/11/16/18) vaccine in males 16 to 26 years old. Clin Vaccine Immunol 2012 Feb;19(2):261-267 [FREE Full text] [doi: 10.1128/CVI.05208-11] [Medline: 22155768]
8. Goldstone SE, Jessen H, Palefsky JM, Giuliano AR, Moreira ED, Vardas E, et al. Quadrivalent HPV vaccine efficacy against disease related to vaccine and non-vaccine HPV types in males. Vaccine 2013 Aug 20;31(37):3849-3855. [doi: 10.1016/j.vaccine.2013.06.057] [Medline: 23831322]
9. Elbasha EH, Dasbach EJ. Impact of vaccinating boys and men against HPV in the United States. Vaccine 2010 Oct 04;28(42):6858-6867. [doi: 10.1016/j.vaccine.2010.08.030] [Medline: 20713101]
10. Marty R, Roze S, Bresse X, Largeron N, Smith-Palmer J. Estimating the clinical benefits of vaccinating boys and girls against HPV-related diseases in Europe. BMC Cancer 2013 Jan 08;13:10 [FREE Full text] [doi: 10.1186/1471-2407-13-10] [Medline: 23298365]
11. Kulasingam S, Connelly L, Conway E, Hocking JS, Myers E, Regan DG, et al. A cost-effectiveness analysis of adding a human papillomavirus vaccine to the Australian National Cervical Cancer Screening Program. Sex Health 2007 Sep;4(3):165-175. [Medline: 17931529]
12. Kim JJ, Andres-Beck B, Goldie SJ. The value of including boys in an HPV vaccination programme: a cost-effectiveness analysis in a low-resource setting. Br J Cancer 2007 Nov 05;97(9):1322-1328 [FREE Full text] [doi: 10.1038/sj.bjc.6604023] [Medline: 17923869]
13. Burger EA, Sy S, Nygård M, Kristiansen IS, Kim JJ. Prevention of HPV-related cancers in Norway: cost-effectiveness of expanding the HPV vaccination program to include pre-adolescent boys. PLoS One 2014;9(3):e89974 [FREE Full text] [doi: 10.1371/journal.pone.0089974] [Medline: 24651645]
14. Pearson AL, Kvizhinadze G, Wilson N, Smith M, Canfell K, Blakely T. Is expanding HPV vaccination programs to include school-aged boys likely to be value-for-money: a cost-utility analysis in a country with an existing school-girl program. BMC Infect Dis 2014 Jun 26;14:351 [FREE Full text] [doi: 10.1186/1471-2334-14-351] [Medline: 24965837]
15. Stanley M. HPV vaccination in boys and men. Hum Vaccin Immunother 2014;10(7):2109-2111 [FREE Full text] [doi: 10.4161/hv.29137] [Medline: 25424825]
16. Bruni L, Barrionuevo-Rosa L, Albero G, Serrano B, Mena M, Gomez D, et al. Human papillomavirus and related disease report in Austria. Summary report 2017.: ICO Information Centre on HPV and Cancer; 2017. URL: http://www.hpvcentre.net/statistics/reports/AUT.pdf [accessed 2017-07-07] [WebCite Cache ID 6rmDMlM8M]
17. Machalek DA, Chow EPF, Garland SM, Wigan R, Cornall AM, Fairley CK, et al. Human papillomavirus prevalence in unvaccinated heterosexual men after a national female vaccination program. J Infect Dis 2017 Jan 15;215(2):202-208. [doi: 10.1093/infdis/jiw530] [Medline: 27815379]
18. Garland SM, Molesworth EG, Machalek DA, Cornall AM, Tabrizi SN. How to best measure the effectiveness of male human papillomavirus vaccine programmes? Clin Microbiol Infect 2015 Sep;21(9):834-841 [FREE Full text] [doi: 10.1016/j.cmi.2015.05.038] [Medline: 26055415]
19. Kavanagh K, Sinka K, Cuschieri K, Love J, Potts A, Pollock KGJ, et al. Estimation of HPV prevalence in young women in Scotland; monitoring of future vaccine impact. BMC Infect Dis 2013 Nov 05;13:519 [FREE Full text] [doi: 10.1186/1471-2334-13-519] [Medline: 24188790]
20. Giuliano AR, Lee J, Fulp W, Villa LL, Lazcano E, Papenfuss MR, et al. Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. Lancet 2011 Mar 12;377(9769):932-940 [FREE Full text] [doi: 10.1016/S0140-6736(10)62342-2] [Medline: 21367446]
21. Nyitray AG, Carvalho DSRJ, Baggio ML, Lu B, Smith D, Abrahamsen M, et al. Age-specific prevalence of and risk factors for anal human papillomavirus (HPV) among men who have sex with women and men who have sex with men: the HPV in men (HIM) study. J Infect Dis 2011 Jan 01;203(1):49-57 [FREE Full text] [doi: 10.1093/infdis/jiq021] [Medline: 21148496]
22. Gargano JW, Unger ER, Liu G, Steinau M, Meites E, Dunne E, et al. Prevalence of genital human papillomavirus in males, United States, 2013-2014. J Infect Dis 2017 Apr 01;215(7):1070-1079. [doi: 10.1093/infdis/jix057] [Medline: 28170037]
23. Prescott TL, Phillips IG, DuBois LZ, Bull SS, Mustanski B, Ybarra ML. Reaching adolescent gay, bisexual, and queer men online: development and refinement of a national recruitment strategy. J Med Internet Res 2016;18(8):e200 [FREE Full text] [doi: 10.2196/jmir.5602] [Medline: 27492781]

24. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
25. Sensis. Sensis social media report 2016: how Australian people and businesses are using social media. 2016. URL: https://www.sensis.com.au/asset/PDFdirectory/Sensis_Social_Media_Report_2016.PDF [accessed 2017-07-07] [WebCite Cache ID 6rmGQs6JO]
26. Subasinghe AK, Nguyen M, Wark JD, Tabrizi SN, Garland SM. Targeted Facebook advertising is a novel and effective method of recruiting participants into a human papillomavirus vaccine effectiveness study. JMIR Res Protoc 2016 Jul 22;5(3):e154 [FREE Full text] [doi: 10.2196/resprot.5679] [Medline: 27450586]
27. Kapp JM, Peters C, Oliver DP. Research recruitment using Facebook advertising: big potential, big challenges. J Cancer Educ 2013 Mar;28(1):134-137. [doi: 10.1007/s13187-012-0443-z] [Medline: 23292877]
28. Mishra GD, Hockey R, Powers J, Loxton D, Tooth L, Rowlands I, et al. Recruitment via the Internet and social networking sites: the 1989-1995 cohort of the Australian Longitudinal Study on Women's Health. J Med Internet Res 2014;16(12):e279 [FREE Full text] [doi: 10.2196/jmir.3788] [Medline: 25514159]
29. Ahmed N, Jayasinghe Y, Wark JD, Fenner Y, Moore EE, Tabrizi SN, et al. Attitudes to Chlamydia screening elicited using the social networking site Facebook for subject recruitment. Sex Health 2013 Jul;10(3):224-228. [doi: 10.1071/SH12198] [Medline: 23702200]
30. Gunasekaran B, Jayasinghe Y, Fenner Y, Moore EE, Wark JD, Fletcher A, et al. Knowledge of human papillomavirus and cervical cancer among young women recruited using a social networking site. Sex Transm Infect 2013 Jun;89(4):327-329. [doi: 10.1136/sextrans-2012-050612] [Medline: 23047877]
31. Miyagi E, Motoki Y, Asai-Sato M, Taguri M, Morita S, Hirahara F, et al. Web-based recruiting for a survey on knowledge and awareness of cervical cancer prevention among young women living in Kanagawa prefecture, Japan. Int J Gynecol Cancer 2014 Sep;24(7):1347-1355 [FREE Full text] [doi: 10.1097/IGC.0000000000000220] [Medline: 25054449]
32. Raviotta JM, Nowalk MP, Lin CJ, Huang H, Zimmerman RK. Using Facebook™ to recruit college-age men for a human papillomavirus vaccine trial. Am J Mens Health 2014 Nov 11:110-119. [doi: 10.1177/1557988314557563] [Medline: 25389213]
33. Pedrana A, Hellard M, Gold J, Ata N, Chang S, Howard S, et al. Queer as F**k: reaching and engaging gay men in sexual health promotion through social networking sites. J Med Internet Res 2013;15(2):e25 [FREE Full text] [doi: 10.2196/jmir.2334] [Medline: 23391459]
34. Nguyen P, Gold J, Pedrana A, Chang S, Howard S, Ilic O, et al. Sexual health promotion on social networking sites: a process evaluation of The FaceSpace Project. J Adolesc Health 2013 Jul;53(1):98-104. [doi: 10.1016/j.jadohealth.2013.02.007] [Medline: 23583509]
35. Osborne SL, Tabrizi SN, Brotherton JML, Cornall AM, Wark JD, Wrede CD, et al. Assessing genital human papillomavirus genoprevalence in young Australian women following the introduction of a national vaccination program. Vaccine 2015 Jan 01;33(1):201-208. [doi: 10.1016/j.vaccine.2014.10.045] [Medline: 25444787]
36. Commonwealth Department of Health and Aged Care. Measuring Remoteness: Accessibility/Remoteness Index of Australia (ARIA) - revised edition 2001. URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/E2EE19FE831F26BFCA257BF0001F3DFA/$File/ocpanew14.pdf [accessed 2017-02-20] [WebCite Cache ID 6rmHmVKnm]
37. Australian Bureau of Statistics. 2011 census data. 2013. URL: http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/quickstat/0 [accessed 2017-07-07] [WebCite Cache ID 6rmI1zoSy]
38. Richters J, Altman D, Badcock PB, Smith AMA, de Visser RO, Grulich AE, et al. Sexual identity, sexual attraction and sexual experience: the Second Australian Study of Health and Relationships. Sex Health 2014 Nov;11(5):451-460. [doi: 10.1071/SH14117] [Medline: 25376998]
39. Martin L, Freedman E, Burton L, Rutter S, Knight V, D'Amato A, et al. The C-project: use of self-collection kits to screen for Chlamydia trachomatis in young people in a community-based health promotion project. Sex Health 2009 Jun;6(2):157-162. [doi: 10.1071/SH08039] [Medline: 19457296]
40. Novak DP, Karlsson RB. Simplifying chlamydia testing: an innovative Chlamydia trachomatis testing approach using the internet and a home sampling strategy: population based study. Sex Transm Infect 2006 Apr;82(2):142-7; discussion 152 [FREE Full text] [doi: 10.1136/sti.2005.016832] [Medline: 16581742]
41. Scholes D, Heidrich FE, Yarbro P, Lindenbaum JE, Marrazzo JM. Population-based outreach for Chlamydia screening in men: results from a randomized trial. Sex Transm Dis 2007 Nov;34(11):837-839. [doi: 10.1097/OLQ.0b013e31805ba860] [Medline: 17538514]
42. van Bergen J, Götz HM, Richardus JH, Hoebe CJPA, Broer J, Coenen AJT, PILOT CT study group. Prevalence of urogenital Chlamydia trachomatis increases significantly with level of urbanisation and suggests targeted screening approaches: results from the first national population based study in the Netherlands. Sex Transm Infect 2005 Feb;81(1):17-23 [FREE Full text] [doi: 10.1136/sti.2004.010173] [Medline: 15681716]
43. Giuliano AR, Palefsky JM, Goldstone S, Moreira ED, Penny ME, Aranda C, et al. Efficacy of quadrivalent HPV vaccine against HPV Infection and disease in males. N Engl J Med 2011 Feb 03;364(5):401-411 [FREE Full text] [doi: 10.1056/NEJMoa0909537] [Medline: 21288094]

44. Goldstone S, Palefsky JM, Giuliano AR, Moreira ED, Aranda C, Jessen H, et al. Prevalence of and risk factors for human papillomavirus (HPV) infection among HIV-seronegative men who have sex with men. J Infect Dis 2011 Jan 01;203(1):66-74 [FREE Full text] [doi: 10.1093/infdis/jiq016] [Medline: 21148498]
45. Zou H, Tabrizi SN, Grulich AE, Hocking JS, Bradshaw CS, Cornall AM, et al. Site-specific human papillomavirus infection in adolescent men who have sex with men (HYPER): an observational cohort study. Lancet Infect Dis 2015 Jan;15(1):65-73. [doi: 10.1016/S1473-3099(14)70994-6] [Medline: 25435055]
46. Royal Australian College of General Practitioners. Guidelines for preventative activities in general practice 2016. URL: http://www.racgp.org.au/your-practice/guidelines/redbook/6-communicable-diseases/62-sexually-transmissible-infections/ [accessed 2017-07-08] [WebCite Cache ID 6rmIklAUo]
47. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young Australian women. Am J Epidemiol 2015 May 15;181(10):737-746. [doi: 10.1093/aje/kwv006] [Medline: 25883155]
48. Vardas E, Giuliano AR, Goldstone S, Palefsky JM, Moreira ED, Penny ME, et al. External genital human papillomavirus prevalence and associated factors among heterosexual men on 5 continents. J Infect Dis 2011 Jan 01;203(1):58-65 [FREE Full text] [doi: 10.1093/infdis/jiq015] [Medline: 21148497]
49. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Recruiting young gay and bisexual men for a human papillomavirus vaccination intervention through social media: the effects of advertisement content. JMIR Public Health Surveill 2017 Jun 02;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]

Abbreviations

EOI: expression of interest
HPV: human papillomavirus
HYM: HPV in Young Males
SNS: social networking site
VACCINE: Vaccine Against Cervical Cancer Impact and Effectiveness

Edited by G Eysenbach; submitted 13.08.17; peer-reviewed by MP Nowalk, C Zhang; comments to author 20.09.17; revised version received 03.10.17; accepted 03.10.17; published 17.11.17.

Please cite as:
Das R, Machalek DA, Molesworth EG, Garland SM
Using Facebook to Recruit Young Australian Men Into a Cross-Sectional Human Papillomavirus Study
J Med Internet Res 2017;19(11):e389
URL: http://www.jmir.org/2017/11/e389/
doi:10.2196/jmir.8739
PMID:29150417

Original Paper

Social Media Interventions to Promote HIV Testing, Linkage, Adherence, and Retention: Systematic Review and Meta-Analysis

Bolin Cao[1,2,3], PhD; Somya Gupta[4], MA; Jiangtao Wang[5], BA; Lisa B Hightow-Weidman[6], MD; Kathryn E Muessig[7], PhD; Weiming Tang[2,3,6], PhD; Stephen Pan[2,3], PhD; Razia Pendse[4], MD; Joseph D Tucker[2,3,6], MD, PhD

[1]School of Media and Communication, Shenzhen University, Shenzhen, China

[2]University of North Carolina Project-China, Guangzhou, China

[3]SESH Global, Guangzhou, China

[4]World Health Organization South-East Asia Region, New Delhi, India

[5]School of Journalism and Communication, Jinan University, Guangzhou, China

[6]School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[7]Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Bolin Cao, PhD
School of Media and Communication
Shenzhen University
No. 3688, Nanhai Avenue, Nanshan District
Shenzhen, 518000
China
Phone: 86 13247393943
Fax: 86 0755 26535374
Email: caobolin@szu.edu.cn

Abstract

Background: Social media is increasingly used to deliver HIV interventions for key populations worldwide. However, little is known about the specific uses and effects of social media on human immunodeficiency virus (HIV) interventions.

Objective: This systematic review examines the effectiveness of social media interventions to promote HIV testing, linkage, adherence, and retention among key populations.

Methods: We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist and Cochrane guidelines for this review and registered it on the International Prospective Register of Systematic Reviews, PROSPERO. We systematically searched six databases and three conference websites using search terms related to HIV, social media, and key populations. We included studies where (1) the intervention was created or implemented on social media platforms, (2) study population included men who have sex with men (MSM), transgender individuals, people who inject drugs (PWID), and/or sex workers, and (3) outcomes included promoting HIV testing, linkage, adherence, and/or retention. Meta-analyses were conducted by Review Manager, version 5.3. Pooled relative risk (RR) and 95% confidence intervals were calculated by random-effects models.

Results: Among 981 manuscripts identified, 26 studies met the inclusion criteria. We found 18 studies from high-income countries, 8 in middle-income countries, and 0 in low-income countries. Eight were randomized controlled trials, and 18 were observational studies. All studies (n=26) included MSM; five studies also included transgender individuals. The focus of 21 studies was HIV testing, four on HIV testing and linkage to care, and one on antiretroviral therapy adherence. Social media interventions were used to do the following: build online interactive communities to encourage HIV testing/adherence (10 studies), provide HIV testing services (9 studies), disseminate HIV information (9 studies), and develop intervention materials (1 study). Of the studies providing HIV self-testing, 16% of participants requested HIV testing kits from social media platforms. Existing social media platforms such as Facebook (n=15) and the gay dating app Grindr (n=10) were used most frequently. Data from four studies show that HIV testing uptake increased after social media interventions (n=1283, RR 1.50, 95% CI 1.28-1.76). In the studies where social media interventions were participatory, HIV testing uptake was higher in the intervention arm than the comparison arm (n=1023, RR 1.64, 95% CI 1.19-2.26).

Conclusions: Social media interventions are effective in promoting HIV testing among MSM in many settings. Social media interventions to improve HIV services beyond HIV testing in low- and middle-income countries and among other key populations need to be considered.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO): CRD42016048073; http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016048073 (Archived by WebCite at http://www.webcitation.org/6usLCJK3v)

(J Med Internet Res 2017;19(11):e394) doi:10.2196/jmir.7997

KEYWORDS

social media; HIV; MSM; intervention; testing; adherence

Introduction

Since 2000, global efforts at human immunodeficiency virus (HIV) control have had significant effects. Acquired immune deficiency syndrome (AIDS)-related deaths have been reduced [1], and 19.5 million people are accessing antiretroviral therapy (ART) [2]. To achieve the end of the AIDS epidemic by 2030, 90% of all people living with HIV (PLWH) should know their HIV status, 90% of all people with diagnosed HIV infections should receive ART, and 90% of all people receiving ART should have viral suppression by 2020 [3]. However, in 2016, only 70% of PLWH globally had been diagnosed, 53% were receiving ART, and 44% had achieved viral suppression [4]. Improved efforts are needed to reach PLWH with a comprehensive package of HIV interventions, including HIV testing, linkage to care, ART, and retention. More programs are needed for key populations who are disproportionately affected by HIV and who have difficulty accessing services across the HIV care continuum [5]. The World Health Organization defines key populations as men who have sex with men (MSM), people who inject drugs (PWID), prisoners, sex workers, and transgender people [5,6]. Barriers to HIV interventions among key populations are complex and include persistent stigma and discrimination, punitive laws, and low risk perception [7-9]. Innovative approaches to reach these populations with equitable, accessible, and acceptable services will be essential to achieve 90-90-90 targets [3].

As social media has expanded globally, these platforms have been adopted to deliver HIV interventions, especially for key populations [10,11]. Social media is defined as an Internet-based platform that allows the creation and exchange of user-generated content, usually using either mobile or Web-based technologies [12]. Popular social media platforms, such as Facebook and YouTube, have over 1.5 billion monthly active users as of June 2017 [13]. Social media possesses the characteristics of interactivity, allows users to generate content [14], and attracts high user engagement [15]. Specifically, social media can enable convenient access, at any time and place, to information and services on stigmatized diseases such as HIV. In addition, social media can be used to form online communities to seek social support, which is known to improve treatment adherence and uptake of HIV services [16].

Systematic reviews have been conducted recently on the relationship between social media and HIV outcomes [10,11,17,18]. These studies generally define social media broadly (eg, to include eHealth and mHealth interventions) or examine usage of social media for a variety of purposes, including recruitment, surveillance, communication, and HIV prevention and treatment. These studies have not focused on social media interventions among key populations [5,6]. Key populations are of special importance because they influence epidemic dynamics and likely have a disproportionate influence on the effectiveness of the response to HIV [5]. To address these gaps in a rapidly growing field, this systematic review and meta-analysis looks at the effectiveness of social media interventions in promoting HIV testing, linkage to care, adherence to treatment, and retention along the HIV care continuum among key populations.

Methods

Conduct of Systematic Review

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [19]. In August 2016, the following databases were systematically searched without restriction on publication date: PubMed, Cochrane Library, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Embase, Scopus, and Sociological Abstracts. Three conference databases (Conference on Retroviruses and Opportunistic Infections; International AIDS Society Conference on HIV Science, and Youth+Tech+Health Conference) were also searched for abstracts for 2015 and 2016.

Search Strategy and Selection Criteria

The search strategy was designed with a librarian to identify studies regarding social media interventions to promote HIV testing, linkage, adherence, and retention among key populations (Multimedia Appendix 1). It was developed based on key terms, medical subject headings (MeSH) terms, synonyms, and subject headings related to three groups: (1) HIV, (2) social media or social media category or social media platform, including the most popular social networks (websites and apps) and top gay dating apps worldwide selected on the basis of their popularity and number of active users [20-22], and (3) key populations, including MSM, PWID, sex workers, and transgender persons.

Following the PRISMA guidelines [19], key study characteristics such as population, intervention, and outcomes defined the eligibility criteria. In particular, studies were included when (1) study population included key populations such as MSM, PWID, sex workers, and transgender individuals, (2) the intervention was created or implemented on at least one

social media platform (existing platform or new platform with social networking components), and (3) outcomes included promoting HIV testing, linkage to care and treatment, adherence to ART, or retention. The inclusion criterion did not have any restrictions on geography or setting, except the language of publication as English. Studies without a comparator arm were also included. Studies in which social media were used for marketing/advertising, surveillance, recruitment, or data collection purposes only were excluded. Commentary, protocol, featured article, published articles without full text or sufficient details on interventions or outcomes, and modeling studies were also excluded.

Two investigators independently reviewed all abstracts identified through searches and screened them for eligibility. The full texts of the abstracts that met the eligibility criteria were then reviewed to confirm inclusion in the analysis. Disagreements were resolved through discussion with a third reviewer.

Data Extraction

Data extraction for the included studies was completed using a standardized extraction form in Microsoft Excel that included the following information: first author, study design (randomized controlled trial [RCT] or observational study), study date, study location, sample size, target population (MSM, PWID, transgender individuals, and/or sex worker), intervention dates/duration, social media platforms, the role of social media, the reach of social media intervention, whether there were interactions in-person or offline events, step in the HIV cascade, and study outcomes. The "role of social media" was categorized into (1) social media used to develop intervention materials for promoting HIV services, (2) social media used to establish virtual peer-mentored or online communities that promote HIV interventions, (3) social media as a platform to offer HIV-related services, such as HIV self-testing kits order and request, and (4) social media as a platform to disseminate HIV-related information [23]. Among various social media interventions, those who used interactive characteristics, tailored contents, or peer influence of social media were considered as participatory social media interventions. For studies with a comparator arm, data were abstracted on the type of interventions, participants, and outcomes for both the intervention and the comparator arm.

Quality Assessment

Two reviewers assessed the quality of the included studies using the checklist tool in Sanderson et al [24], and a third reviewer collated the results. For each study, the following six domains were used to assess risk of bias: (1) methods for selecting study participants, (2) methods for measuring exposure and outcome variables, (3) design-specific source of bias, (4) method of control confounding, (5) statistical methods, and (6) other biases (including conflict of interest and disclosure of funding sources). The Cochrane Collaboration's recommendations [25] were used to categorize each of the six domains as "low risk of bias" ("+") or "high risk of bias" ("–").

Statistical Analysis

For RCTs meeting the inclusion criteria, pooled relative risks (RR) were used to compare the participants in the intervention and the comparator arm with respect to HIV testing rates among total participants. Pooled RR were also used to compare HIV testing outcomes at baseline and post social media interventions. Data from studies on HIV self-sampling were pooled to summarize (1) the proportion of total participants requesting HIV self-sampling services, (2) the proportion of participants requesting HIV self-sampling who returned their test kits, and (3) HIV positivity rates. Meta-analyses were conducted by Review Manager, version 5.3. Pooled RR and 95% confidence intervals (CI) were calculated using random-effects models.

Results

Study Characteristics

Of the 981 articles and abstracts, 26 studies met our inclusion criteria [26-51] (Figure 1). Of the 26 studies, eight were RCTs [28,33,41,42,46,47,49,50] and 18 were observational studies [26,27,29-32,34-40,43-45,48,51] (Table 1). These studies were implemented in 10 countries from 2007-2015 and published from 2011-2016. In total, 18 studies were from four high-income countries as defined by the World Bank [28,29,31-38,40,42-45,47,48,50], 10 studies were from the United States [28,33-35,38,42,43,45,47,50], five from the United Kingdom [29,31,32,36,48], two each from Australia [40,44], China [46,51], and Thailand [26,27], and one each from Taiwan [37], India [41], Peru [49], Mexico [30], and Guatemala [39]. The reach of social media interventions varied from 55 to over 17,000 individuals.

Key Populations

MSM were the primary population included in the reviewed social media interventions. All of the included studies contained MSM populations, and five also covered transgender individuals [26,27,35,42,46]. In three studies, MSM with the following specific characteristics were included: young MSM [28], young black MSM [47], and MSM living with HIV [33].

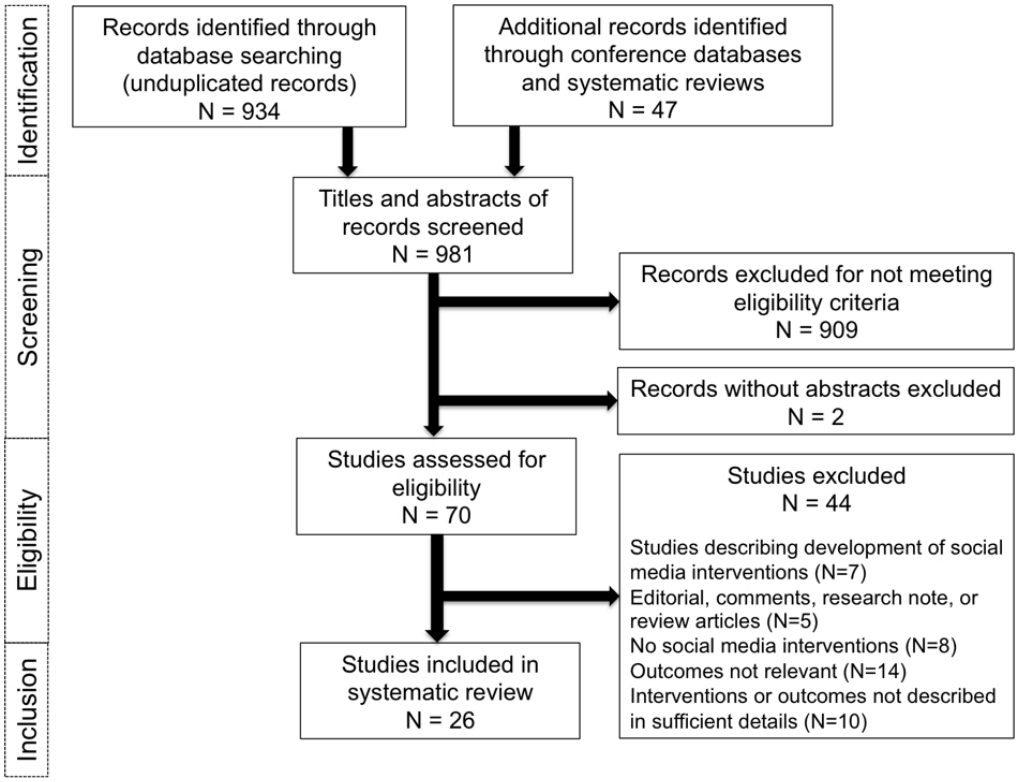
Social Media Platforms

Existing social and sexual networking sites and gay-specific websites/apps were used for interventions. Fifteen studies used Facebook [26-29,32,33,35-37,39,41,44,47,49,50] and 10 studies used Grindr, a social networking app catering to MSM [29,32,34-36,38-40,45,48]. Other generic social networking platforms included YouTube, Twitter, and QQ, and other social networking sites or apps catering to MSM included Gaydar, Jack'd, Scruff, A4A, and Radar. Six studies used both social and sexual networking sites as intervention platforms [29,32,35,36,39,51]. Five studies created their own social media platform to provide HIV services [26-28,30,33].

Role of Social Media

In one study, social media was used as a crowdsourcing tool to develop intervention materials (video) for promotion of HIV testing [46]. The study showed that the crowdsourced video arm (114/307, 37.1%) had similar results as the social marketing arm (111/317, 35.0%) in promoting HIV tests, but the cost of crowdsourced intervention was less than the social marketing intervention per first-time HIV test (US $131 vs US $238 per person) and per new HIV diagnosis (US$415 vs US $799 per person).

Figure 1. PRISMA chart showing study selection process.



Ten studies reported using social media to establish virtual communities where HIV services (testing and adherence) were promoted [26,33,37,40,42,43,45,47,49,50]. While two studies built their own platforms to form interactive communities [26,33], the remaining studies established these communities on Facebook, popular gay dating apps or existing chat rooms [37,40,42,43,45,47,49,50]. Interventions provided in the communities included online HIV counseling, educating people about HIV and importance of HIV testing, referral for HIV testing, and answering questions related to HIV. HIV testing uptake was higher in virtual communities that emphasized peer-to-peer interaction among participants [37,49].

Social media also served as a platform for delivering services such as HIV self-testing kits, HIV self-sampling, or home-based HIV testing (9 studies) [27,29,31,32,34,36,44,48,50]. MSM were often recruited through a broadcast message as a pop-up or a personal message or promotional banner on social media, and the most popular platforms were Facebook and Grindr [29,32,34]. Finally, HIV-related information was disseminated using social media in nine studies [28-30,35,38,39,41,44,51]. Some studies tailored the content to be distributed on social media by user's race/ethnicity and age based on quantitative data and evaluated through community advisory boards and youth advisory boards [28]. Other studies trained outreach health educators on using gay-specific apps to provide MSM with HIV education, counseling, and testing information [38,39].

In three studies, social media played more than one role [29,44,50]. In 14 studies, the online activities that were a part of the social media interventions were augmented with offline events or in-person interaction [26,27,29-32,35,36,38,39,44,48,49,51]. For example, during or post social media interventions, HIV testing were provided at clinics or home [27,36,38,51] and in studies on HIV test home-sampling, test results were conveyed to participants over the phone or via messages [29].

Table 1. Characteristics of the studies included in the systematic review (n=26).

Study	Study design	Study location	Sample size	Target population	Intervention period/ duration	Social media platforms	Role of social media	Interaction in-person or offline	Step in the HIV cascade
Anand 2015 [26]	Observational	Thailand	11,120	MSM & transgender individuals	2011.09-2014.12 (ongoing)	Facebook, Line, "Adam's Love" online platform	Build community[a]	Yes	HIV testing and linkage to care
Anand 2016 [27]	Observational	Thailand	97	MSM & transgender individuals	2015.12-2016.05	Facebook, Line, "Adam's Love" online platform	Deliver kits[b]	Yes	HIV testing
Bauermeister 2015 [28]	RCT	US	130	Young MSM	30 days	Facebook; "Get connected" platform	Disseminate[c]	No	HIV testing
Brady 2014 [29]	Observational	UK	17,629	MSM	Phase 1: 2013.01-2013.09; Phase 2: 2013.11-2014.03	Grindr, Gaydar, Manhunt, Facebook, Twitter	Deliver kits (blood) & disseminate	Yes	HIV testing and linkage to care
Buzdugan 2016 [30]	Observational	Mexico	61	MSM	5 weeks	Online game app	Disseminate	Yes	HIV testing
Elliot 2012 [31]	Observational	UK	321	MSM	2011.11.07-2012.01.11	Gaydar	Deliver kits	Yes	HIV testing
Elliot 2016 [32]	Observational	UK	17,361	MSM	2012-2014	Gaydar, Grindr, Facebook	Deliver kits (oral)	Yes	HIV testing and linkage to care
Horvath 2013 [33]	RCT	US	123	MSM living with HIV	2011.02-2011.04	Facebook; "Thrive with me" platform	Build community	No	ART adherence
Huang 2015 [34]	Observational	US	16,328 (112 in survey)	MSM	Phase 1: 2014.04.17-2014.05.29; Phase 2: 2014.10.13-2014.11.11	Grindr	Deliver kits (oral)	No	HIV testing
Hyden 2016 [35]	Observational study	US	274	MSM & transgender individuals	2013.03-2017.12 (ongoing)	Facebook, Scruff, Grindr, Online game App	Disseminate	Yes	HIV testing
Jones 2015 [36]	Observational	UK	305	MSM	2014.11.29-2014.11.30 (2 days)	Facebook, Grindr, Squirt	Deliver kits (blood)	Yes	HIV testing
Ko 2013 [37]	Observational	Taiwan	1037	MSM	2011.04-2011.09	Facebook	Build community	No	HIV testing
Lampkin 2016 [38]	Observational	US	903	MSM	2012.10-2013.03; 2013.10-2014.03	Grindr	Disseminate	Yes	HIV testing
Mendizabal-Burastero 2016 [39]	Observational	Guatemala	7244	MSM	2014.07-2015.12	Facebook, Twitter, Grindr, WhatsApp	Disseminate	Yes	HIV testing and linkage to care
Munro 2016 [40]	Observational	Australia		MSM	Not mentioned	Grindr	Build community	No	HIV testing
Patel 2016 [41]	RCT	India	244	MSM	2015.02-2015.05	Facebook, WhatsApp	Disseminate	No	HIV testing

Study	Study design	Study location	Sample size	Target population	Intervention period/ duration	Social media platforms	Role of social media	Interaction in-person or offline	Step in the HIV cascade
Rhodes 2016 [42]	RCT	US	1292	MSM & transgender individuals	2013.07-2014.06	Adam4Adam, BlackGayChat, Craigslist, Gay.com	Build community	No	HIV testing
Rhodes 2011 [43]	Observational	US	509	MSM	2009.02-2009.07	Online chat room	Build community	No	HIV testing
Roberts 2015 [44]	Observational	Australia	503	MSM	2013.11 & 2014.07	Facebook, YouTube, Twitter	Deliver kits & disseminate	Yes	HIV testing
Sun 2015 [45]	Observational	US	2709	MSM	2013.08-2014.02	A4A Radar, Grindr, Jack'd, Scruff	Build community	No	HIV testing
Tang 2016 [46]	RCT	China	721	MSM & transgender individuals	2014.09	MSM dating websites	Develop materials[d]	No	HIV testing
Washington 2016 [47]	RCT	US	142	Young black MSM	Not mentioned	Facebook	Build community	No	HIV testing
West 2015 [48]	Observational	UK	55	MSM	2014.11 (5 days)	Grindr	Deliver kits	Yes	HIV testing
Young 2015 [49]	RCT	Peru	556	MSM	Phase 1: 2012.03.19-2012.06.11; Phase 2: 2012.09.26-2012.12.19	Facebook	Build community	Yes	HIV testing
Young 2013 [50]	RCT	US	112	MSM	2010.09-2011.01	Facebook	Build community & deliver kits	No	HIV testing
Zou 2013 [51]	Observational	China	429	MSM	2007.06-2007.08	Gay dating & volunteer websites, QQ, online gay chatroom	Disseminate	Yes	HIV testing

[a]"Build community" is when social media is used to establish virtual peer-mentored or online communities that promote HIV interventions.

[b]"Deliver kits" is when social media serves as a platform to deliver newly designed or extant evidence-based HIV intervention.

[c]"Disseminate" is when social media is used as a platform to disseminate HIV-related information.

[d]"Develop materials" is when social media is used to develop intervention materials for promoting HIV services.

Uptake of HIV Testing and HIV Home Sampling

We found 25 studies that used social media interventions to promote HIV testing [26-32,34-51] among MSM or MSM and transgender individuals, while one study used it for improving ART adherence [33]. Of these 25 studies, seven were RCTs [27,40,41,45,46,48,49] and the remaining 18 were observational studies [26,27,29-32,34-40,43-45,48,51]. Four studies were pooled to compare the rate of HIV testing at baseline and postintervention [37,41-43]. The meta-analysis showed that the HIV testing rate significantly increased after the social media interventions were provided (RR=1.50, 95% CI 1.28-1.76, I^2 66%; Figure 2).

Data from five RCTs show that when social media interventions were participatory [28,42,46,49,50], HIV testing rates were significantly higher (RR 1.64, 95% CI 1.19-2.26, I^2 75%; Figure 3) than in the comparative arm, where there were no social media interventions, or the social media interventions provided general health information or were not participatory. Of the seven RCTs, Patel et al [41] was excluded since both arms were interventional arms while Washington et al [47] was excluded from the analysis due to nonavailability of complete data. Similar to included studies, in Washington et al [47], those in the intervention arm (HIV testing video + chat) were seven times more likely to test for HIV than those in control group receiving standard HIV text information (study not included in pooled RR due to incomplete data).

Four studies in the United Kingdom and United States used social media to offer HIV self-testing services to MSM [29,31,32,50] (Tables 2,3, and 4). On average, 15.65% of the participants requested HIV testing kits (n=67,054; 95% CI 15.37-15.92), ranging from 15.50% to 36.36% (data from three studies) [30,31,49]. Nearly 57% of the participants who

requested test kits returned them (n=24,703; 95% CI 55.92-57.16). On average, the HIV positivity rate was 1.51% among those who received results (n=13,956; 95% CI 1.32-1.73) (data from three studies [29,31,32,34]).

Quality Assessment

There was a high risk of bias in methods for selecting study participants in 15 studies [27,29-32,34-36,38-40,42-44,48] and in measuring exposure and outcome bias in 13 studies [26,29,31,34-36,38-40,44,45,48,51]. Design-specific source of bias was present in 10 studies [27,28,30,36,39,41,45,46,49,50]. Concerns regarding control confounding bias were in 14 studies [26,27,29-32,34,35,37,40,44,45,48,51] and concerns of statistical methods in 15 studies [26,29-32,34-36,38-40,44,48,49,51] (Multimedia Appendix 2). Only two studies had low risk of bias in all the 6 dimensions [33,47]. See Multimedia Appendix 3 for results of the included studies.

Figure 2. Comparison between studies with baseline and postintervention data on HIV testing (n=4).

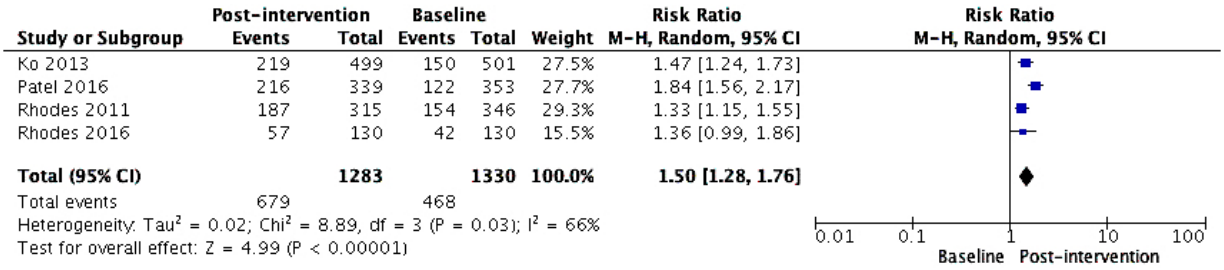


Figure 3. Comparison of HIV testing uptake between the intervention arm and the comparison arm in the RCTs (n=5).

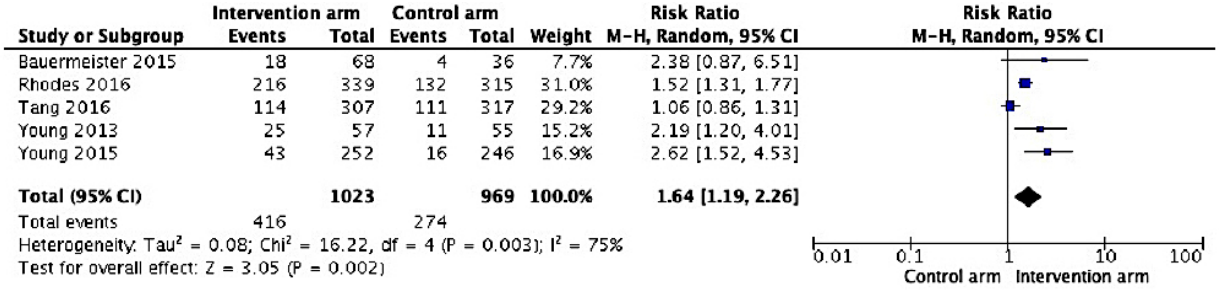


Table 2. Studies that use social media to provide HIV self-testing services (n=3): Request for HIV test kits.

Study	Total participants/webpage views	Request testing kit	Request percentage (95% CI)
Elliot 2012 [31]	363	132	36.36 (31.58-41.43)
Elliot 2016 [32]	66,579	10,323	15.50 (15.23-15.78)
Young 2013 [50]	112	36	32.14 (24.21-41.26)
Total	67,054	10,491	15.65 (15.37-15.92)

Table 3. Studies that use social media to provide HIV self-testing services (n=4): Return rate of test kits.

Study	Request for testing kit	Returned testing kit	Return percentage (95% CI)
Brady 2014 [29]	14,212	8187	57.61 (56.79-58.42)
Elliot 2012 [31]	132	73	55.30 (46.79-63.52)
Elliot 2016 [32]	10,323	5696	55.18 (54.22-56.13)
Young 2013 [50]	36	11	30.56 (18.00-46.86)
Total	24,703	13,967	56.54 (55.92-57.16)

Table 4. Studies that use social media to provide HIV self-testing services (n=4): HIV positivity rate.

Study	Total participants tested	HIV positive	Positive rate (95% CI)
Brady 2014 [29]	8187	111	1.36 (1.13-1.63)
Elliot 2012 [31]	73	4	5.48 (2.15-13.26)
Elliot 2016 [32]	5696	96	1.69 (1.38-2.05)
Total	13,956	211	1.51 (1.32-1.73)

Discussion

Principal Findings

Social media have played varied roles and have been used in multiple settings to promote public health interventions. For example, social media was found to be effective in promoting physical activity [52] and smoking cessation [53]. Previous reviews have also found that social media interventions can improve HIV-related outcomes such as promoting HIV testing and linkage to care [10,11,17,18]. This systematic review shows that social media interventions are being used to promote HIV testing among MSM and transgender populations. In the studies included, HIV testing uptake among this key population group increased after implementation of social media interventions. Also, HIV testing uptake was particularly higher when the social media interventions were participatory and peer-driven. These findings are consistent with other studies [54,55] and support implementation of social media interventions that reach MSM populations and promote MSM participation.

HIV testing is the main focus of social media interventions along the HIV care continuum and is critical for achieving the first Joint United Nations Programme on HIV/AIDS (UNAIDS) 90-90-90 target [3]. Four studies found that HIV self-testing kits distributed through social media led to large numbers of requests for and return of test kits. World Health Organization guidelines strongly recommend HIV self-testing as an additional approach to HIV testing services [56]. One study on HIV self-testing included in this review concludes that providing self-testing through social media is acceptable and cost-effective [34]. Cost per person tested was US $39 and US $26 of this cost was attributed to test kits [57]. This is comparable to other research that found self-testing (US $9.23) costs lower than facility-based testing (US $11.84) in Malawi [58]. The current studies that use social media to reach hard-to-reach populations and offer HIV self-sampling or self-testing services were implemented in high-income countries where social media penetration rates are high. More usability research on HIV self-testing using social media is needed in low- and middle-income countries. Meanwhile, all included articles reported using social media to promote HIV testing, except that one study aimed to enhance ART adherence [33]. In addition to promoting HIV testing, future research should explore how social media can be leveraged to promote other HIV care services.

Social media interventions that were participatory and peer-led resulted in higher HIV testing rates compared with those that had no social media interventions or social media interventions that did not include interactive characteristics. Engaging people to participate in the interventions can increase their perceived relevance and importance of the activities [59]. Participatory strategies of social media interventions, such as tailoring intervention contents to user's characteristics [28,49] and encouraging peer interactions [49,50] and community contributions [46], can increase the likelihood for participants to link to HIV care and services. Provided that participatory and peer-led social media interventions allow the organizers to receive feedback from the users and can motivate users to take actions, future social media interventions should adopt established strategies or develop new strategies to improve engagement and interaction with participants.

Among the studies included in this review, a variety of social media platforms were used for promoting HIV services. They included existing online social and sexual networking platforms and standalone HIV-focused websites or apps. Over half of the studies used Facebook or one of the globally or nationally popular gay dating apps. Existing platforms may be useful for HIV interventions as they often have wide reach, high user engagement and retention, and attract specific key populations that would otherwise be difficult to recruit [60]. On the other hand, five of the 26 studies developed standalone platforms. Such platforms could be used to design more tailored and innovative interventions, but using only standalone platforms may have the drawback of poor retention and limited coverage [61]. Few studies used other popular social media platforms such as Twitter and YouTube, partly because such platforms are often used as knowledge, news, or video sharing platforms but not to develop long-term relationships for continuous service provision [62]. Research is needed to assess the appropriateness and effectiveness of these platforms to deliver HIV-specific interventions for key populations.

All studies in this review primarily focused on MSM populations. This may be due to the early adoption and popularity of social and sexual networking sites among this population. Online sex seeking sites and apps for MSM provide an important channel for seeking social support and meeting new sexual partners [63,64]. While use of these platforms has been associated with increased risk behaviors for HIV [65,66], they have also been effectively used for promotion of HIV testing [47,67]. Although several studies included transgender people in addition to MSM, results were not disaggregated and studies were not powered to evaluate differences in transgender populations. The feasibility and acceptability of social media interventions and the appropriate social media platforms to deliver HIV services within networks of PWID and sex workers is also underresearched. Given disparities along the prevention and care continuum with these populations [68] and strong within-group networks [69], the potential of using social media in these populations should be actively explored.

Limitations

This systematic review found that studies on social media intervention have several limitations. First, the current research's methodological scoring was low due to the fact that the majority of the studies were exploratory and descriptive. Of the 26 studies, 24 had high or unclear risk of bias for at least one of the bias items in the methodological quality assessment. The high risk of bias indicates relatively low quality of studies on social media interventions, hindering more pooled analysis and causing possible bias of this meta-analysis. Second, there was also a variety of control groups used in the seven RCTs on HIV testing. For example, in the control groups, either no interventions were provided or social media was used to provide general health-related information or promote HIV services without interactive features and peer-leaders. Third, some studies used apps to improve linkage and retention in care, but for

one-way monitoring of patients instead of bidirectional interactions.

This systematic review also has its own limitations. First, it was restricted to English language publications and was not able to cover studies in other languages. However, an empirical analysis found that this was not associated with bias [70]. Second, due to the heterogeneity of intervention types and measures in outcomes, we had difficulties pooling certain data. Some pooled results were largely driven by a few studies. This meta-analysis was not able to compare effects of social media intervention on MSM versus transgender individuals and through standalone platform versus existing platforms. Last, currently many studies on this topic are underway, so the state-of-the-science will continue to grow rapidly over the next few years.

Future Studies

This systematic review shows the effectiveness of using social media interventions to improve HIV testing among MSM population and has implications for both future research and public policy. Gaps should be filled on using social media to promote HIV services beyond HIV testing and among key populations beyond MSM. Moreover, how to maximize the use of social media to promote HIV service needs further exploration. In addition, given that social media interventions have been found to be effective, there is an opportunity for national programs to leverage social media to support scale-up of such interventions. In particular, low- and middle-income countries, where mobile usage is rising, seem to have relatively few social media interventions.

Conclusion

A total of 26 articles were identified in this systematic review to examine the role of rapidly expanding social media in improving access to HIV-related interventions and its effect in promoting HIV services among key populations. Social media can contribute to creating innovative intervention programs, disseminating intervention information, building virtual communities, and especially promoting HIV self-testing and self-sampling. Social media interventions were effective in increasing HIV testing rates, especially in high-income countries, aimed at MSM. To achieve the goal of 90-90-90 by 2020 and ultimately end AIDS by 2030, adapting these social media interventions in low- to middle-income countries and other key populations may be useful.

Acknowledgments

Support of this work was provided by the National Institutes of Health (National Institute of Allergy and Infectious Diseases 1R01AI114310); UNC-South China STD Research Training Centre (Fogarty International Centre 1D43TW009532); UNC Center for AIDS Research (National Institute of Allergy and Infectious Diseases 5P30AI050410); National Social Science Foundation of China (16CXW031); and the Bill & Melinda Gates Foundation to the MeSH Consortium (BMGF-OPP1120138). This publication was also supported by Grant Number UL1TR001111 from the National Center for Advancing Translational Sciences (NCATS) at the National Institutes of Health.

We acknowledge Jennifer S Walker, UNC, a Chapel Hill librarian, for conducting the initial search for studies, and Gabriella Stein, undergraduate student, UNC, Chapel Hill, for her support at the early stage of this study.

The findings and conclusions in this study are those of the authors and do not necessarily represent the views of the World Health Organization.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms.

[PDF File (Adobe PDF File), 24KB - jmir_v19i11e394_app1.pdf]

Multimedia Appendix 2

Quality assessment of the included studies.

[PDF File (Adobe PDF File), 34KB - jmir_v19i11e394_app2.pdf]

Multimedia Appendix 3

Results of included studies.

[PDF File (Adobe PDF File), 34KB - jmir_v19i11e394_app3.pdf]

References

1. Joint UN Programme on HIV/AIDS (UNAIDS). Prevention Gap Report. Geneva, Switzerland URL: http://www.unaids.org/sites/default/files/media_asset/2016-prevention-gap-report_en.pdf [accessed 2017-11-05] [WebCite Cache ID 6ukNT13cH]
2. UNAIDS. UNAIDS announces 18.2 million people on antiretroviral therapy, but warns that 15-24 years of age is a highly dangerous time for young women. URL: http://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2016/november/20161121_PR_get-on-the-fast-track [accessed 2017-09-16] [WebCite Cache ID 6tX2aDSGD]
3. Joint UN Programme on HIV/AIDS (UNAIDS). 90-90-90 - An ambitious treatment target to help end the AIDS epidemic. 2017 Jan 01. URL: http://www.unaids.org/en/resources/documents/2017/90-90-90 [accessed 2017-08-18] [WebCite Cache ID 6snZis7u6]
4. UNAIDS. Ending AIDS: Progress towards the 90-90-90 targets. 2017. URL: http://www.unaids.org/sites/default/files/media_asset/Global_AIDS_update_2017_en.pdf [accessed 2017-09-18] [WebCite Cache ID 6tZLC1E1o]
5. World Health Organization. Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations. 2014. URL: http://apps.who.int/iris/bitstream/10665/128048/1/9789241507431_eng.pdf [accessed 2017-09-18] [WebCite Cache ID 6tZLPyEtE]
6. World Health Organization. WHO Definition of key terms. 2013 Jun. URL: http://www.who.int/hiv/pub/guidelines/arv2013/intro/keyterms/en/ [accessed 2017-08-11] [WebCite Cache ID 6tZLMqK3o]
7. Altman D, Aggleton P, Williams M, Kong T, Reddy V, Harrad D, et al. Men who have sex with men: stigma and discrimination. Lancet 2012 Jul 28;380(9839):439-445. [doi: 10.1016/S0140-6736(12)60920-9] [Medline: 22819652]
8. Decker MR, Crago A, Chu SKH, Sherman SG, Seshu MS, Buthelezi K, et al. Human rights violations against sex workers: burden and effect on HIV. Lancet 2015 Jan 10;385(9963):186-199 [FREE Full text] [doi: 10.1016/S0140-6736(14)60800-X] [Medline: 25059943]
9. Kamarulzaman A, Altice FL. Challenges in managing HIV in people who use drugs. Curr Opin Infect Dis 2015 Feb;28(1):10-16 [FREE Full text] [doi: 10.1097/QCO.0000000000000125] [Medline: 25490106]
10. Garett R, Smith J, Young SD. A Review of Social Media Technologies Across the Global HIV Care Continuum. Curr Opin Psychol 2016 Jun 01;9:56-66 [FREE Full text] [doi: 10.1016/j.copsyc.2015.10.024] [Medline: 26925455]
11. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman IM. Social Media and HIV: A Systematic Review of Uses of Social Media in HIV Communication. J Med Internet Res 2015 Nov 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
12. Margetts H, John P, Hale S, Yasseri T. How social media shape collective action. New Jersey, US: Princeton University Press; 2015.
13. Constine J. TechCrunch. Facebook now has 2 billion monthly users... and responsibility URL: https://techcrunch.com/2017/06/27/facebook-2-billion-users/ [accessed 2017-08-18] [WebCite Cache ID 6snaB5jFi]
14. Obar JA, Wildman SS. Social media definition and the governance challenge: An introduction to the special issue. Telecommunications Policy 2015;39(9):745-750. [doi: 10.1016/j.telpol.2015.07.014]
15. Saberi P, Johnson M. Correlation of Internet Use for Health Care Engagement Purposes and HIV Clinical Outcomes Among HIV-Positive Individuals Using Online Social Media. J Health Commun 2015;20(9):1026-1032 [FREE Full text] [doi: 10.1080/10810730.2015.1018617] [Medline: 26120890]
16. Mo P, Coulson N. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. Cyberpsychol Behav 2008 Jun;11(3):371-374. [doi: 10.1089/cpb.2007.0118] [Medline: 18537512]
17. Ghosh D, Krishnan A, Gibson B, Brown S, Latkin CA, Altice FL. Social Network Strategies to Address HIV Prevention and Treatment Continuum of Care Among At-risk and HIV-infected Substance Users: A Systematic Scoping Review. AIDS Behav 2017 Apr;21(4):1183-1207. [doi: 10.1007/s10461-016-1413-y] [Medline: 27125244]
18. Muessig K, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman L. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
19. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
20. Statista. 2017. Global active usage penetration of leading social networks as of February 2017 URL: https://www.statista.com/statistics/274773/global-penetration-of-selected-social-media-sites/ [accessed 2017-09-09] [WebCite Cache ID 6tZMIiuiq]
21. Digital Trends. 2017. Find love or have a fling with the best LGBT dating apps for iPhone and Android URL: https://www.digitaltrends.com/mobile/best-lgbt-dating-apps-iphone-android/ [accessed 2017-09-19] [WebCite Cache ID 6tZMQm4go]
22. Johnson N. Plugin Group. 2016. The top 15 most popular social media sites in 2016 URL: https://plugingroup.com/top-15-popular-social-media-sites-november-2016/ [accessed 2017-09-19] [WebCite Cache ID 6tZMVB2xp]
23. Tso L, Tang W, Li H, Yan H, Tucker J. Social media interventions to prevent HIV: A review of interventions and methodological considerations. Curr Opin Psychol 2016 Jun 01;9:6-10 [FREE Full text] [doi: 10.1016/j.copsyc.2015.09.019] [Medline: 26516632]

24. Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol 2007 Jun;36(3):666-676. [doi: 10.1093/ije/dym018] [Medline: 17470488]
25. The Cochrane Collaboration. 2011 Mar. Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0) URL: http://handbook-5-1.cochrane.org/ [accessed 2017-11-06] [WebCite Cache ID 6ukR01O78]
26. Anand T, Nitpolprasert C, Ananworanich J, Pakam C, Nonenoy S, Jantarapakde J, et al. Innovative strategies using communications technologies to engage gay men and other men who have sex with men into early HIV testing and treatment in Thailand. J Virus Erad 2015 Apr 01;1(2):111-115 [FREE Full text] [Medline: 27482400]
27. Anand T, Nitpolprasert C, Kerr S, Apornpong T, Ananworanich J, Phanuphak P, et al. Factors influencing and associated with the decision to join in Thailand's first online supervised HIV self-testing and counselling initiative. J Int AIDS Soc 2016;19((8Suppl 7)):21487. [doi: 10.7448/IAS.19.8.21487]
28. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, Harper GW, Horvath K, Weiss G, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874 [FREE Full text] [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
29. Brady M, Nardone A, Buenaventura E, Qureshi F, Edwardes D, Kelly P. Home HIV sampling linked to national HIV testing campaigns: A novel approach to improve HIV diagnosis. HIV Med 2014;15:7-8.
30. Buzdugan R. Evaluation of a pilot game-based intervention to prevent HIV among young men who have sex with men in Mexico City. 2016 Presented at: Youth + Tech + Heal Conference; 2016; San Francisco, CA.
31. Elliot E, McCowan A, McCormak S, Rossi M, Van Every T. O20 Home sampling through social network websites: can we reduce undiagnosed HIV? Sex Transm Infect 2012 May 23;88(Suppl 1):A8. [doi: 10.1136/sextrans-2012-050601a.20]
32. Elliot E, Rossi M, McCormack S, McOwan A. Identifying undiagnosed HIV in men who have sex with men (MSM) by offering HIV home sampling via online gay social media: a service evaluation. Sex Transm Infect 2016 Sep;92(6):470-473. [doi: 10.1136/sextrans-2015-052090] [Medline: 27221475]
33. Horvath K, Oakes JM, Rosser BRS, Danilenko G, Vezina H, Amico KR, et al. Feasibility, acceptability and preliminary efficacy of an online peer-to-peer social support ART adherence intervention. AIDS Behav 2013 Jul;17(6):2031-2044 [FREE Full text] [doi: 10.1007/s10461-013-0469-1] [Medline: 23553347]
34. Huang E, Marlin RW, Young SD, Medline A, Klausner JD. Using Grindr, a Smartphone Social-Networking Application, to Increase HIV Self-Testing Among Black and Latino Men Who Have Sex With Men in Los Angeles, 2014. AIDS Educ Prev 2016 Aug;28(4):341-350. [doi: 10.1521/aeap.2016.28.4.341] [Medline: 27427928]
35. Hyden C. “Heads or Tails?” Online interactive graphic novel for HIV/STI prevention: Content development and opportunities for adaptation. 2015 Presented at: Youth+Health+Tech Conference; 2015; San Francisco, CA.
36. Jones M, Aderogba K, Drunis L. Providing rapid HIV testing in their homes for men who have sex with men, recruited via social media. HIV Medicine 2015;16:50-51.
37. Ko N, Hsieh C, Wang M, Lee C, Chen C, Chung A, et al. Effects of Internet popular opinion leaders (iPOL) among Internet-using men who have sex with men. J Med Internet Res 2013 Feb 25;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]
38. Lampkin D, Crawley A, Lopez TP, Mejia CM, Yuen W, Levy V. Reaching Suburban Men Who Have Sex With Men for STD and HIV Services Through Online Social Networking Outreach: A Public Health Approach. J Acquir Immune Defic Syndr 2016 May 01;72(1):73-78. [doi: 10.1097/QAI.0000000000000930] [Medline: 27097365]
39. Mendizabal-Burastero R. Reaching the unreachable: MSM recruitment strategy using social networks to HIV prevention services in Guatemala city. 2016 Presented at: International AIDS Society AIDS Conference; 2016; Durban, South Africa.
40. Munro T, Rabiem L, Gray J. Accessing gay men: transitioning from physical venues to online spaces. 2016 Presented at: International AIDS Society AIDS Conference; 2016; Durban, South Africa.
41. Patel V. CHALO! A social media based peer-delivered intervention increases HIV testing in men who have sex with men in Mumbai, India: a randomized trial. 2016 Presented at: International AIDS Society AIDS Conference; 2016; Durban, South Africa.
42. Rhodes SD, McCoy TP, Tanner AE, Stowers J, Bachmann LH, Nguyen AL, et al. Using Social Media to Increase HIV Testing Among Gay and Bisexual Men, Other Men Who Have Sex With Men, and Transgender Persons: Outcomes From a Randomized Community Trial. Clin Infect Dis 2016 Dec 01;62(11):1450-1453 [FREE Full text] [doi: 10.1093/cid/ciw127] [Medline: 26980878]
43. Rhodes S, Vissman A, Stowers J, Miller C, McCoy T, Hergenrather K, et al. A CBPR partnership increases HIV testing among men who have sex with men (MSM): outcome findings from a pilot test of the CyBER/testing internet intervention. Health Educ Behav 2011 Jun;38(3):311-320 [FREE Full text] [doi: 10.1177/1090198110379572] [Medline: 21393625]
44. Roberts N, Holden J, Duck T, Kitchener S. Health promotion 'on steroids': the value of an experiential approach to promote rapid HIV testing in NSW, Australia. Public Health Res Pract 2015 Mar 30;25(2):e2521522 [FREE Full text] [doi: 10.17061/phrp2521522] [Medline: 25848740]
45. Sun CJ, Stowers J, Miller C, Bachmann LH, Rhodes SD. Acceptability and feasibility of using established geosocial and sexual networking mobile applications to promote HIV and STD testing among men who have sex with men. AIDS Behav 2015 Mar;19(3):543-552 [FREE Full text] [doi: 10.1007/s10461-014-0942-5] [Medline: 25381563]

46. Tang W, Han L, Best J, Zhang Y, Mollan K, Kim J, et al. Crowdsourcing HIV Test Promotion Videos: A Noninferiority Randomized Controlled Trial in China. Clin Infect Dis 2016 Dec 01;62(11):1436-1442 [FREE Full text] [doi: 10.1093/cid/ciw171] [Medline: 27129465]
47. Washington T, Applewhite S. Using social media as a platform to direct young Black men who have sex with men to an intervention to increase HIV testing. 2016 Presented at: International AIDS Society AIDS Conference; 2016; Durban, South Africa.
48. West R, Okecha E, Forbes K. Keeping “app” to date: using geolocation apps to signpost to local sexual health services. Sex Transm Infect 2015 Jun 18:A32-A33. [doi: 10.1136/sextrans-2015-052126.96]
49. Young S, Cumberland W, Nianogo R, Menacho L, Galea J, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015 Jan;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
50. Young SD, Cumberland WG, Lee S, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. Ann Intern Med 2013 Sep 03;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]
51. Zou H, Wu Z, Yu J, Li M, Ablimit M, Li F, et al. Internet-facilitated, voluntary counseling and testing (VCT) clinic-based HIV testing among men who have sex with men in China. PLoS One 2013;8(2):e51919 [FREE Full text] [doi: 10.1371/journal.pone.0051919] [Medline: 23418417]
52. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
53. Duke J, Hansen H, Kim A, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. J Med Internet Res 2014 Jul 10;16(7):e169 [FREE Full text] [doi: 10.2196/jmir.3430] [Medline: 25014311]
54. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
55. Noar S, Pierce L, Black H. Can computer-mediated interventions change theoretical mediators of safer sex? A meta-analysis. Hum Commun Res 2010;36(3):261-297. [doi: 10.1111/j.1468-2958.2010.01376.x]
56. World Health Organization. 2016. Guidelines on HIV self-testing and partner notification URL: http://apps.who.int/iris/bitstream/10665/251655/1/9789241549868-eng.pdf?ua=1 [accessed 2017-09-18] [WebCite Cache ID 6tZNkhl0a]
57. Huang E, Marlin R, Medline A, Young S, Daniels J. Cost-effectiveness of HIV self-testing promotion through Grindr, a smartphone social networking application. Sex Transm Infect 2015;91(Suppl 2):A226.
58. Maheswaran H, Petrou S, MacPherson P, Choko AT, Kumwenda F, Lalloo DG, et al. Cost and quality of life analysis of HIV self-testing and facility-based HIV testing and counselling in Blantyre, Malawi. BMC Med 2016 Feb 19;14:34 [FREE Full text] [doi: 10.1186/s12916-016-0577-7] [Medline: 26891969]
59. Petty RE, Cacioppo JT. Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. Journal of Personality and Social Psychology 1979;37(10):1915-1926. [doi: 10.1037/0022-3514.37.10.1915]
60. Hightow-Weidman LB, Muessig KE. New media challenges and opportunities. Sex Transm Infect 2017 Aug;93(5):309-310. [doi: 10.1136/sextrans-2016-052673] [Medline: 27888207]
61. Maher CA, Lewis LK, Ferrar K, Marshall S, De BI, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
62. Lee EJ, Oh SY. Seek and You Shall Find? How Need for Orientation Moderates Knowledge Gain from Twitter Use. J Commun 2013 Jun 24;63(4):745-765. [doi: 10.1111/jcom.12041]
63. 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]
64. Mustanski B, Greene GJ, Ryan D, Whitton SW. Feasibility, acceptability, and initial efficacy of an online sexual health promotion program for LGBT youth: the Queer Sex Ed intervention. J Sex Res 2015;52(2):220-230. [doi: 10.1080/00224499.2013.867924] [Medline: 24588408]
65. Benotsch EG, Kalichman S, Cage M. Men who have met sex partners via the Internet: prevalence, predictors, and implications for HIV prevention. Arch Sex Behav 2002 Apr;31(2):177-183. [Medline: 11974643]
66. McFarlane M, Bull SS, Rietmeijer CA. The Internet as a newly emerging risk environment for sexually transmitted diseases. JAMA 2000 Jul 26;284(4):443-446. [Medline: 10904506]
67. Cao B, Liu C, Durvasula M, Tang W, Pan S, Saffer A, et al. Social Media Engagement and HIV Testing Among Men Who Have Sex With Men in China: A Nationwide Cross-Sectional Survey. J Med Internet Res 2017 Jul 19;19(7):e251 [FREE Full text] [doi: 10.2196/jmir.7251] [Medline: 28724510]
68. Risher K, Mayer KH, Beyrer C. HIV treatment cascade in MSM, people who inject drugs, and sex workers. Curr Opin HIV AIDS 2015 Nov;10(6):420-429 [FREE Full text] [Medline: 26352393]

69. Tucker J, Peng H, Wang K, Chang H, Zhang S, Yang L, et al. Female sex worker social networks and STI/HIV prevention in South China. PLoS One 2011;6(9):e24816 [FREE Full text] [doi: 10.1371/journal.pone.0024816] [Medline: 21931856]
70. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care 2012 Apr;28(2):138-144. [doi: 10.1017/S0266462312000086] [Medline: 22559755]

Abbreviations

AIDS: acquired immune deficiency syndrome
ART: antiretroviral therapy
CINAHL: Cumulative Index to Nursing and Allied Health Literature
HIV: human immunodeficiency virus
MSM: men who have sex with men
PLWH: people living with HIV
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PWID: people who inject drugs
RCT: randomized controlled trial
RR: relative risk
UNAIDS: The Joint United Nations Programme on HIV/AIDS
UNC: University of North Carolina

Edited by G Eysenbach; submitted 08.05.17; peer-reviewed by TED Yeo, C Khosropour, C Mejia, J Andriesen; comments to author 29.07.17; revised version received 19.09.17; accepted 11.10.17; published 24.11.17.

Please cite as:
Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, Pan S, Pendse R, Tucker JD
Social Media Interventions to Promote HIV Testing, Linkage, Adherence, and Retention: Systematic Review and Meta-Analysis
J Med Internet Res 2017;19(11):e394
URL: http://www.jmir.org/2017/11/e394/
doi:10.2196/jmir.7997
PMID:29175811

Original Paper

Two Novel Cognitive Behavioral Therapy–Based Mobile Apps for Agoraphobia: Randomized Controlled Trial

Marina Christoforou[1], BSc (Hons), MSc; José Andrés Sáez Fonseca[1,2], MSc, MD; Elias Tsakanikos[2], BSc (Hons), MSc, PhD

[1]Division of Psychiatry, University College London, London, United Kingdom

[2]Department of Psychology, University of Roehampton, London, United Kingdom

Corresponding Author:
Elias Tsakanikos, BSc (Hons), MSc, PhD
Department of Psychology
University of Roehampton
Holybourne Avenue
London, SW15 4JD
United Kingdom
Phone: 44 020 8392 3080
Email: elias.tsakanikos@roehampton.ac.uk

Abstract

Background: Despite the large body of literature demonstrating the effectiveness of cognitive behavioral treatments for agoraphobia, many patients remain untreated because of various barriers to treatment. Web-based and mobile-based interventions targeting agoraphobia may provide a solution to this problem, but there is a lack of research investigating the efficacy of such interventions.

Objective: The objective of our study was to evaluate for the first time the effectiveness of a self-guided mobile-based intervention primarily targeting agoraphobic symptoms, with respect to a generic mobile app targeting anxiety.

Methods: A Web-based randomized controlled trial (RCT) compared a novel mobile app designed to target agoraphobia (called Agoraphobia Free) with a mobile app designed to help with symptoms of anxiety in general (called Stress Free). Both interventions were based on established cognitive behavioral principles. We recruited participants (N=170) who self-identified as having agoraphobia and assessed them online at baseline, midpoint, and end point (posttreatment) over a period of 12 weeks. The primary outcome was symptom severity measured by the Panic and Agoraphobia Scale.

Results: Both groups had statistically significant improvements in symptom severity over time (difference –5.97, 95% CI –8.49 to –3.44, P<.001 for Agoraphobia Free and –6.35, 95% CI –8.82 to –3.87, P<.001 for Stress Free), but there were no significant between-group differences on the primary outcome (difference 0.38, 95% CI –1.96 to 3.20, P=.64).

Conclusions: This is, to our knowledge, the first RCT to provide evidence that people who identify as having agoraphobia may equally benefit from a diagnosis-specific and a transdiagnostic mobile-based intervention. We also discuss clinical and research implications for the development and dissemination of mobile mental health apps.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 98453199; http://www.isrctn.com/ISRCTN98453199 (Archived by WebCite at http://www.webcitation.org/6uR5vsdZw)

(J Med Internet Res 2017;19(11):e398) doi:10.2196/jmir.7747

KEYWORDS

agoraphobia; anxiety; eHealth; computerized interventions; mobile applications; randomized controlled trial; RCT

Introduction

People with agoraphobia have considerable impairment in their daily lives as a result of persisting avoidance of places and situations. Lifetime prevalence of agoraphobia with and without panic disorder in the adult population has been reported to range from 0.8% to 2.6% in American and European community samples [1,2]. Agoraphobia has been ranked as one of the most chronically persistent disorders [3], with higher rates of work-related inactivity and disability than with other chronic conditions [4,5].

The standard psychological treatment for agoraphobia symptoms is cognitive behavioral therapy (CBT), which is also recommended by the UK National Institute of Health and Care Excellence [6]. People with agoraphobia have catastrophic ideas regarding the likelihood of threat in a situation or environment, such as fainting and being ignored or ridiculed by others [7,8], and as a consequence engage in "safety behaviors" to prevent the expected catastrophe and reduce anxiety. Such safety behaviors preserve and might even enhance the maladaptive beliefs and thus maintain anxiety [9,10]. In CBT, cognitive restructuring is used to challenge catastrophic cognitions and unrealistic predictions, and to generate alternative, more realistic expectations. Behavioral experiments may be conducted to help disconfirm such beliefs and drop safety behaviors [8]. Moreover, imaginal or in vivo exposure to the feared situations is used to reduce situational avoidance and phobic anxiety. Breathing and relaxation exercises are sometimes employed to help the patient cope with overwhelming stress.

The effectiveness of CBT in the treatment of agoraphobia has been demonstrated in several studies. A systematic review and meta-analysis [11] examined the effectiveness of psychotherapy and pharmacotherapy in the treatment of panic disorder, with the majority of participants also having agoraphobia. They identified 23 randomized controlled trials (RCTs) and showed that, although combined psychotherapy and pharmacotherapy produced the best results in the short term, in the long term combined treatment was as effective as psychotherapy alone, and both treatment groups were superior to pharmacotherapy alone. Among the different types of psychotherapy, CBT had the strongest evidence. More recent studies have also provided support for CBT as an effective treatment in agoraphobia patients [12-15] in line with earlier studies [16].

Even though CBT has been established as the preferred and most effective treatment for agoraphobia and panic, many people remain untreated. From the population of people with anxiety disorders in developed countries, only 16.7% seek help from a mental health professional, and only 21.3% of those receive CBT [17]. Despite the fact that CBT has nowadays become more widely available, certain barriers to treatment may explain the low levels of self-seeking in people with agoraphobia. For example, fear of stigmatization, lack of psychoeducation, long waiting lists, transportation problems, and time constraints significantly limit access to CBT [18-20]. In addition, the very nature of agoraphobia, which may include fears of leaving the house and using public transport, may make it even more difficult for people to actively seek professional help. To overcome such barriers, effective interventions that are easily accessible and do not require therapist face-to-face contact should be available.

Internet-based or computerized interventions can be considered as acceptable alternatives to standard treatments, as they can be clinically effective and minimize treatment barriers for users. Such interventions can be self- or therapist guided, presenting materials of cognitive behavioral principles and methods in a series of lessons, which are typically accompanied by homework tasks and supplementary information. Systematic reviews show that computerized CBT (CCBT) is as beneficial as therapist-led CBT in the treatment of anxiety disorders such as panic disorder, and is more effective than treatment as usual or waiting list conditions [21,22]. CCBT can also reduce therapist time [21] and have good acceptability, as shown by adherence and satisfaction levels [22]. A more recent review of Internet CBT, including 8 trials of panic disorder with and without agoraphobia, concluded that Internet CBT was as efficacious as face-to-face CBT and more efficacious than waiting list, attention, information, and online discussion control groups [23].

One promising mode of delivering computerized interventions are mobile phones, because of their relatively low cost and widespread use [24]. Since users carry their mobile devices with them in almost any situation, mobile phones, and particularly smartphones, might facilitate engagement with exposure exercises in the users' natural environments. Although mobile apps have been tested for several conditions, such as unipolar depression [25], borderline personality disorder, and substance abuse [26], to date there has been no study testing an app that primarily targets agoraphobic symptoms. Because of the isolative nature of agoraphobia and the extreme avoidance behaviors, an app that could be easily downloaded over the Internet onto patients' phones or tablets, requiring no traveling to sites, might be especially appealing and convenient for this population.

This study was a Web-based RCT aiming to test the clinical effectiveness of a novel mobile app for agoraphobia in a community-based sample. The treatment app, Agoraphobia Free, developed by Health eLiving Partnership Ltd (HeLP) for the iOS and Android operating systems, provides an interactive game-based intervention using cognitive behavioral techniques that target agoraphobia and panic. The comparator was a stress-reduction app (Stress Free), which does not address agoraphobic symptoms or panic, but stress and anxiety in general. Both interventions were self-guided and were evaluated over a period of 12 weeks. Adults that self-identified with agoraphobia were randomly allocated to the 2 treatment arms and completed self-reported assessments at baseline, midpoint (6 weeks), and end point (12 weeks) of the trial. The primary objective was to examine whether an agoraphobia-specific intervention would be more effective than a generic, anxiety-related intervention. A secondary aim was also to assess the level of engagement with these interventions and the feasibility of conducting such a trial over the Web.

Methods

Design

The study was a Web-based, assessor-blinded, parallel-group RCT with an active control group. Participants were individually randomly allocated (ratio of 1:1) to either the treatment group or the control group at baseline and were given an equal amount of time to complete each intervention (12 weeks). Data were gathered automatically through online collection of anonymized data, without any researcher intervention. The primary outcome was the degree of symptom severity, as measured by the self-reported version of the Panic and Agoraphobia Scale [27] at end point.

Participants

Inclusion Criteria

Participants needed to be adults (aged ≥18 years) and identify themselves as having agoraphobia. Participants also had to be willing and able to provide informed consent to participate in the trial.

We used no diagnostic check, as the aim was to recruit a community sample that would reflect the nature of the population that would use the apps in a real-world setting, where no screening or check would be required.

Exclusion Criteria

The exclusion criteria were as follows: (1) inability to give informed consent due to significant cognitive or intellectual impairment, (2) no adequate understanding of English as a first language, and (3) not having a mobile device than could run the app as designed.

Recruitment and Setting

A website was set up for the trial on which advertisements and all relevant information and updates were posted. The website was hosted by HeLP Ltd (currently known as Thrive Therapeutic Software). Advertisements of the trial were also posted on social media (eg, Facebook, Twitter), support groups and forums, websites of relevant organizations such as Anxiety UK, blogs, and university websites. We also individually contacted members of anxiety support groups via Facebook or forum messages. We created a mailing list of people subscribing their interest in the trial, whom we encouraged to stay in touch until the trial commenced. Moreover, contacts of the Chief Investigator that were working in health-related settings were encouraged to inform any relevant clients or representatives about the study. There were 2 rounds of recruitment in order to achieve a larger number of participants.

We always contacted participants via email. Initially, those who expressed interest in participating followed a link to the online information sheet and consent form, which outlined the eligibility criteria and information about the trial. Through consenting to participate and answering a series of questions, participants confirmed that they met the criteria and understood the purpose of the study. Participants also provided their email addresses and names, though the latter was optional. Participant codes were then assigned to those who consented, and details about their mobile devices (smartphones or tablets) and demographics (age and sex) were obtained. Participants were instructed to use only their participant code to identify themselves.

Participants received invitations for the apps, which were available to use for free. Emails also contained links to weekly surveys on app use and links to questionnaires, a description of the specific survey questionnaire, links to the calendar and the main website of the trial, information about the upcoming survey, and useful contact details. A Frequently Asked Questions section was set up on the main website. Data were collected online and could only be accessed by the researchers.

Interventions

Agoraphobia Free

The treatment app was Agoraphobia Free (version 0.8), developed by Thrive Therapeutic Software for the treatment of agoraphobia, and this was the first time that it was evaluated. The app was a game-based interactive intervention, with 3-dimensional characters and situations that simulate real-life environments. Specifically, the app presented a case example of a virtual character who had agoraphobia. The user was required to guide her, through the help of the virtual therapist, to complete the different therapeutic tasks. Those tasks were based on CBT principles, comprising psychoeducation, reflection, cognitive restructuring, interoceptive exposure, and systematic desensitization. The 3 overarching goals were to decrease the virtual character's catastrophic cognitions, safety behaviors, and physiological arousal. In this way, users were able to progress through the character's recovery and treatment, and build the formulation of her difficulties. At the same time, users were asked to apply the techniques they used in the case example to their own situation. Therefore, by using the character's recovery journey as a template, users learned how to reduce their catastrophic thinking, their safety behaviors, and physiological arousal.

At the beginning of the intervention, we asked participants to set a hierarchy of goals they would like to achieve (eg, going out of the house, going to the supermarket). At the end of each session, they were required to complete each one of those goals in order, using the techniques and strategies they had learned in the intervention. The sessions were designed so that the tasks became increasingly more challenging as participants progressed through the intervention. The sessions needed to be completed in the order they were presented in for the next ones to become unlocked. There were 10 sessions in total, and the participants were asked to complete 1 or 2 sessions per week at their own pace, and reminders for those were sent weekly. Multimedia Appendix 1 shows an outline of each session.

Stress Free

The control app was Stress Free (version 1.3) developed by Thrive Therapeutic Software, which focused on teaching relaxation techniques and generic CBT skills though a virtual therapist. The app also included a few distraction techniques presented in the form of games that required attention to help individuals cope with acute anxiety. The intervention was presented initially as a linear training program using video and audio guides. The user first learned diaphragmatic breathing, then differential deep muscle relaxation, then self-hypnosis, and finally meditation. These relaxation techniques have been previously shown to be effective in reducing stress [28,29]. Participants rated their anxiety before and after relaxation sessions using a visual analog scale. After completing the training, they were familiarized with CBT concepts such as negative automatic thoughts and the process of challenging them. Finally, we gave participants a daily goal to complete, such as doing 3 sessions of diaphragmatic breathing or a 10-minute session of meditation. In between sessions, participants were prompted to record their anxiety on a CBT journal. The app only targeted stress and anxiety in general, so

it was not specific to agoraphobia. In the original version, there was no maximum limit of sessions that users could go through until they mastered the techniques. However, for the purpose of the trial, we asked participants to complete 10 sessions in total, so as to match the number of sessions required from the treatment group.

Table 1 compares the 2 apps by showing which components and exercises were present in each. Although some components were common to both apps, in Agoraphobia Free they were specifically tailored to agoraphobia. In Stress Free the exercises addressed stress and anxiety in general without referring to agoraphobia. The 2 apps were matched for the number of sessions required and time to complete the interventions (minimum: 6 weeks, maximum: 12 weeks). No training, supervision, or guidance was offered before or during the trial, and only a basic description of each app was provided. Any questions participants had regarding the app or any technical issues they encountered were resolved through email. Weekly reminders and short surveys were sent to promote engagement and monitor progress, and the completion of 1 to 2 sessions per week was recommended for both groups. We also informed participants at the beginning of the trial that, when they completed the intervention, they would receive a link to download the app they did not receive for free, as a reward for taking part and an additional incentive to complete their assigned intervention. Participants were assured that they would not be asked any questions about the second app, as it would not be part of the research.

The 2 apps were available on Android and iOS. Multimedia Appendix 2 shows screenshots of the apps.

Outcomes

The primary outcome was the severity of agoraphobic and panic symptoms, measured by the PAS [27]. The questionnaire was administered online in a self-report format at baseline, midpoint (6 weeks), and end point (12 weeks) of the trial. Reminders were emailed to those who did not reply to the questionnaires before the prespecified deadline. Participants rated the symptoms they experienced in the previous week on a 5-point scale. The questionnaire comprises 14 items, although only 13 of those are used to calculate severity scores. The items are grouped into 5 subscales: (1) Panic Attacks, assessing frequency, severity, and duration of panic attacks, (2) Agoraphobic Avoidance, assessing frequency of avoidance, and number and relevance of avoided situations, (3) Anticipatory Anxiety, assessing frequency and severity of anxiety, (4) Disability, assessing impairment in family life, social relationships, and employment, and (5) Worries about Health, assessing worries about damage to health and assumption of organic disease. The scale was originally validated in a sample of 235 panic patients and has shown good internal consistency (Cronbach alpha=.88), test-retest reliability, and good internal and external validity [27,30]. Cronbach alphas for this study's sample indicated good internal consistency, with alpha=.84 for the overall scale. Coefficients for individual subscales ranged from .41 to .88. The scale has been shown to be sensitive to change due to treatment in 2 clinical trials [31,32].

The secondary outcomes were completion of the intervention and engagement with the apps. Completion of the interventions was assessed in the short online surveys that were sent weekly, by asking participants if they had used the app, how much time they used it for over the past week, and how many sessions they had completed. If participants claimed that they had not used the app, they were asked to give reasons.

Sample Size

We estimated the sample size on the basis of using the self-rated version of the PAS (SD 10.3) as the primary outcome measure. At least 68 participants in each arm were needed to detect a 5-point between-group difference, with a 2-sided significance level set at 5% and power at 80%. Given the high dropout rates in Web-based trials [33], the aim was to recruit at least 150 participants in total.

Table 1. The different features present (indicated by "X") in each app.

App features	Stress Free app	Agoraphobia Free app
Relaxation technique training	X	X
Automated activity goals	X	X
CCBT[a] basic tutorial	X	X
CCBT journal with prompts	X	X
Maintenance sessions	X	X
Self-soothing strategies	X	X
Distraction techniques	X	X
Structured CCBT program		X
Goal setting by user		X
Construction of exposure hierarchy		X
Development of a formulation		X
Relapse prevention session		X

[a]CCBT: computerized cognitive behavioral therapy.

Randomization

We used a random computer-generated sequence to randomly allocate participants to the 2 intervention groups. The random allocation sequence was retrieved from a website that generates truly random numbers [34], by a person outside the research team. We applied block randomization to ensure equal numbers of participants in each group (ratio of 1:1). Participants were automatically allocated to intervention groups by a formula on Excel version 14.6.6 (Microsoft Corporation) using the random number sequence, which was coordinated by another contact who was not a member of the research team. The random sequence and the allocation of participants to groups were concealed from research staff throughout the trial. The same person sent emails to participants containing the link to the assigned app after they had returned the baseline questionnaire.

Blinding

The trial was assessor blinded, as researchers were blinded to treatment allocation throughout the trial and during the statistical analysis. This was achieved by having a person outside the research team to manage treatment allocation and personal communications with the participants. Any questions or comments made in the surveys were forwarded from this contact to research staff, excluding any participant details or codes. This was to ensure that the researchers remained blinded to treatment allocation, as some comments contained information about the specific app the participants were using. Researchers did not have any access to data regarding treatment allocation, as those were stored on a secure database, separate and protected from other research files. During data collection and analysis, only numerical codes were used to indicate group allocation. Researchers did not know which groups those referred to until the end of the analysis.

Blinding the participants was not possible, as the apps were clearly labelled as “Stress Free” and “Agoraphobia Free,” and masking those would require significant changes in the software, which were not feasible. Moreover, even if the apps were not differentially labelled, it is very likely that participants would have become aware of which treatment group they were assigned to because of the intervention content and the extent to which it addressed agoraphobia.

Ethics

The research was approved by the Roehampton University Ethics Committee (reference number: PSYC 14/ 117). The trial was registered and reported in accordance with the Consolidated Standards of Reporting Trials (CONSORT)-EHEALTH checklist (Multimedia Appendix 3) [35]. Data were kept anonymized and protected according to the UK Data Protection Act [36]. We could obtain information about adverse events or effects of the interventions from feedback participants provided in the weekly surveys.

Analysis

We analyzed the data on Stata version 14 (StataCorp LLC). We checked baseline data for normality and obtained descriptive statistics to capture the demographic and clinical characteristics of the sample. The analysis performed was intention-to-treat, as requested in the CONSORT [37]. All participants who completed the baseline assessment were included in the analysis as they were randomly allocated.

We used a linear mixed model to analyze the data, with a random effect of participant, and fixed effects of time (baseline, midpoint, end point), group (Agoraphobia Free and Stress Free), and the interaction between time and group. The estimated baseline PAS score was constrained to be identical in the 2 groups, thus adjusting for baseline and allowing the relationship between baseline and follow-up scores to differ at each time point. Another advantage of this statistical method is that the data from all participants contribute to the analysis, even if there is a substantial amount of missing data at follow-up [38]. We used an unstructured residual covariance matrix to allow for correlations within participants between the different time points. Statistical significance was taken at the 5% level ($P<.05$).

We conducted a planned secondary completers’ analysis using the same data analytic strategy as the intention-to-treat analysis. This analysis included only those participants who were identified as intervention completers.

Results

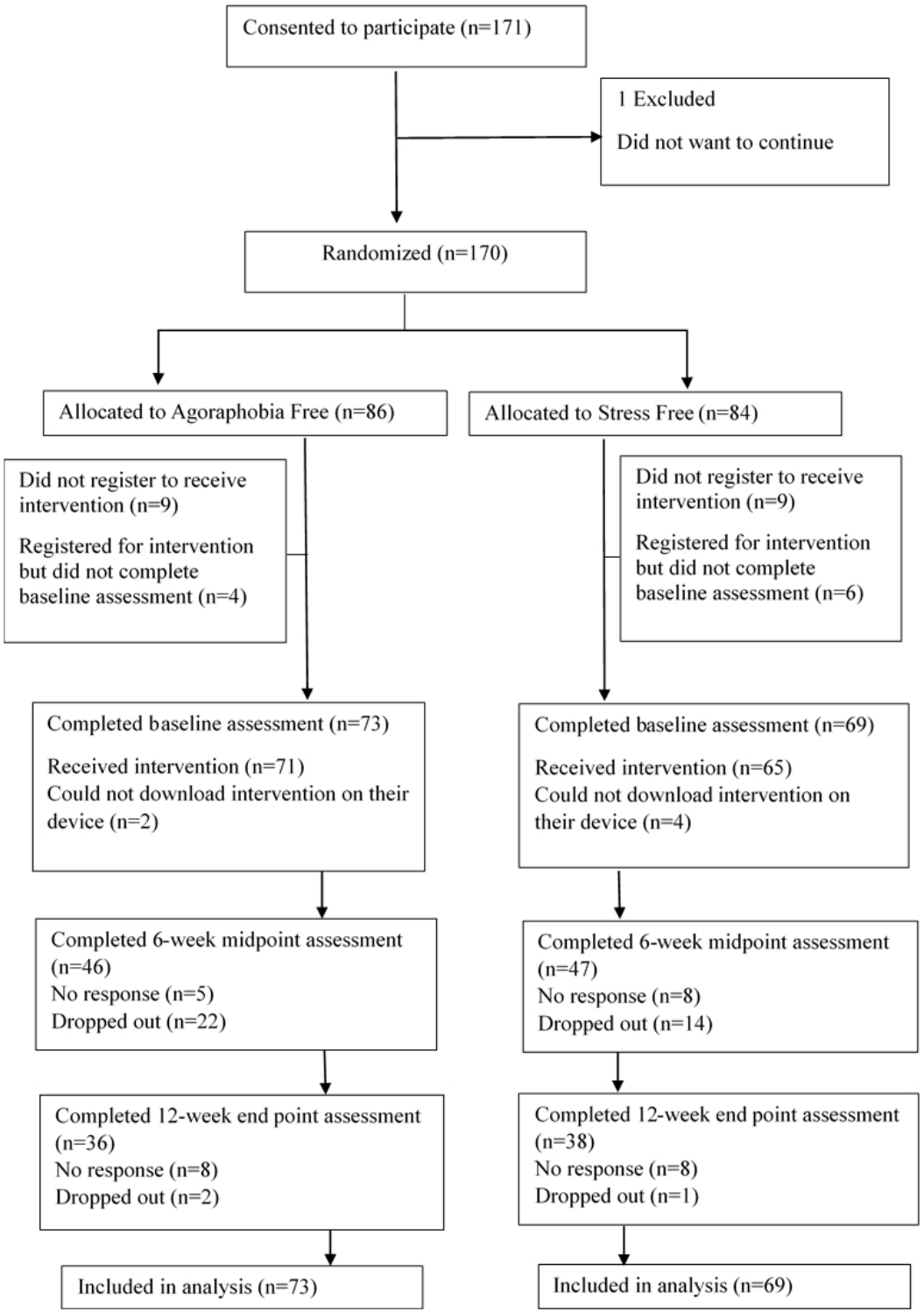
Participant Flow

The first phase of recruitment started from September 2014 and ended in late February 2015, and the second phase started in March 2015 and concluded in April, 2015. In the first round, 153 participants consented to participate, and we recruited 17 additional participants in the second round. The procedure following recruitment was the same in both samples. Data collection ended in June 2015. Figure 1 shows how the total sample of 170 participants progressed through the trial.

After 171 individuals consented to participate, 1 wanted to withdraw from the trial. The rest were randomly allocated into the 2 arms, but were not told which app they were assigned to at that point. Although the aim was to assign equal numbers of participants to each group, at each stage of randomization, the treatment group happened to have 1 more person than the other group, because the number of participants recruited each time was odd, and researchers were blinded to the allocation sequence.

We asked participants to fill in a form with their demographic and device details to register for the intervention so that we could set up the corresponding app invitations. A total of 18 participants did not return the form and did not reply to emails, even after they had been sent several reminders; thus, we treated them as dropouts and excluded them from the trial. The baseline PAS questionnaire was then sent and had to be completed in order for participants to proceed and receive the intervention. A total of 10 individuals did not return the questionnaire and did not reply to any emails. Up to this point, participants were still unaware of the group they were assigned to; thus, their exclusion from the trial was very unlikely to introduce any bias. The remaining 142 who completed the PAS were sent their allocated intervention, but 6 participants could not download the app on their device, despite efforts to resolve the technical issues.

Figure 1. Flow of participants through the different stages of the trial.



After baseline, 39 participants did not want to continue with the trial and dropped out, while 29 participants did not reply to emails and did not complete the questionnaires. We included all participants who provided baseline data in the analysis.

Baseline Characteristics

Of the 142 participants who completed the baseline assessment, 2 did not give details of their age and sex (1 from each treatment arm). Of the remaining 140, 118 (84.3%) were female and had a mean age of 39.7 years (SD 11.3). Overall, the mean participant PAS total score was 30.3 (SD 8.7), and the scores were normally distributed (Shapiro-Wilk test: W=.985, P=.14).

The mean PAS total score fells into the severe range and was much higher than that of the original sample used in the validation of the scale [27], which reported a mean score of 23.5, SD 10.3. Our sample also had a higher proportion of women than the original sample (of whom only 57% were women) and a slightly higher age (mean 36.09 years in the original sample). A significant proportion of our sample (114/142, 80.3%) had severe or extremely severe Agoraphobic

Avoidance, while 85.2% (121/142) avoided more than 8 situations. None of the participants scored 0 on the Agoraphobic Avoidance subscale, which indicates that all of them experienced agoraphobic symptoms to some degree. The lowest score on that scale was 1.67, indicating mild Agoraphobic Avoidance. In contrast, 29 participants (20.4%) obtained a score of 0 on the Panic Attacks subscale, which suggests that some participants experienced agoraphobia in the absence of panic attacks. Overall, agoraphobic symptoms were more prominent than panic symptoms in this sample.

Table 2 shows the demographic and clinical baseline characteristics of the participants by group. The 2 groups did not differ statistically on any of those characteristics at baseline (all $P>.05$).

Missing Data

A total of 68 (47.9%) participants had missing outcomes. Overall, the differences between participants with missing data and those without were not statistically significant on any of the baseline variables examined. There were no significant differences in age ($t_{138}=0.85$, $P=.40$) sex ($\chi^2_1=0.6$, $P=.45$), or clinical symptom severity ($t_{140}=1.32$, $P=.19$). There were 37 participants (54%) with missing data in the Agoraphobia Free arm and 31 (46%) in the Stress Free arm. The relationship between missing data and treatment arm was not significant ($\chi^2_1=0.5$, $P=.49$). Therefore, participant attrition did not seem to be biased with regard to group or any other baseline factor.

Main Analysis

We produced a linear mixed model assessing the relative effects of each intervention on PAS scores at the 2 follow-up time points. Table 3 presents the estimated differences in PAS scores for the Agoraphobia Free group compared with the Stress Free group adjusted for baseline score at the 2 time points.

At end point, symptom severity scores decreased in both groups, but there was no evidence that the changes were significantly greater among participants of the Agoraphobia Free group than among those in the Stress Free group. Similarly, at midpoint there were no significant differences in symptom severity changes between the 2 groups. Therefore, there were no significant differences between the 2 groups. Figure 2 presents the differences on the primary outcome for each group over time.

We carried out the same linear mixed model analysis (n=142) using each PAS subscale as the dependent variable to examine whether there was a difference between the 2 groups in terms of symptom dimensions. We found no significant interactions between group and time for any of those outcomes (all $P>.05$).

We conducted within-group contrasts to examine the degree of change in symptom severity over time. For both Agoraphobia Free (n=73) and Stress Free (n=69), there were statistically significant improvements in symptom severity from baseline to midpoint and end point. Table 4 presents the results.

Table 2. Baseline demographic and clinical characteristics of each treatment arm.

Characteristics	Agoraphobia Free app (n=73)	Range of PAS[a] scores	Stress Free app (n=69)	Range of PAS scores
Age (years), mean (SD)	39.21 (10.45)	N/A[b]	40.23 (12.21)	N/A
Sex (female), n (%)	64 (88.9)	N/A	54 (79.4)	N/A
PAS scores, mean (SD)				
Total	30.77 (8.72)	9-50	29.80 (8.72)	6-47
Panic Attacks	1.58 (1.05)	0-4	1.52 (0.88)	0-3.33
Agoraphobic Avoidance	3.30 (0.57)	1.67-4	3.24 (0.52)	2-4
Anticipatory Anxiety	2.70 (0.94)	0-4	2.59 (1.01)	0-4
Disability	2.39 (1.16)	0-4	2.47 (1.09)	0-4
Worries about Health	1.78 (1.17)	0-4	1.46 (1.11)	0-4

[a]PAS: Panic and Agoraphobia Scale.

[b]N/A: not applicable.

Table 3. Intention-to-treat analysis at end point (12 weeks) and midpoint (6 weeks), n=142 .

Time point of PAS[a] score	Agoraphobia Free mean (SD)	Stress Free mean (SD)	Effect estimate		
			Difference	95% CI	*P* value
End point (primary outcome)	24.33 (16.81)	23.95 (16.51)	0.38	–1.96 to 3.20	.64
Midpoint	27.66 (13.37)	27.03 (13.27)	0.62	–3.13 to 3.89	.83

[a]PAS: Panic and Agoraphobia Scale.

Figure 2. Clinical symptom severity as indicated by the total score on the Panic and Agoraphobia Scale (PAS) in each group at trial baseline, midpoint, and end point.

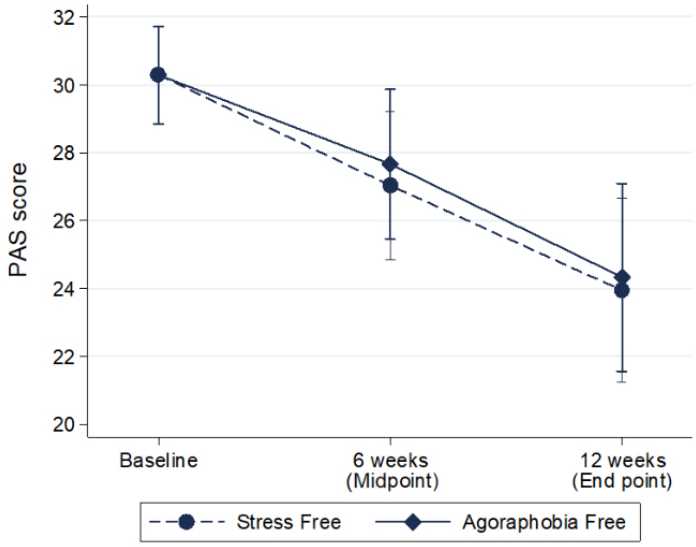


Table 4. Within-group contrasts capturing the differences in Panic and Agoraphobia Scale total score between time points within each treatment arm (n=142).

Treatment arms	Baseline-midpoint contrast			Baseline-end point contrast		
	Difference	95% CI	*P* value	Difference	95% CI	*P* value
AF[a]	–2.64	–4.48 to –0.79	.005	–5.97	–8.49 to –3.44	<.001
SF[b]	–3.25	–5.09 to –1.43	<.001	–6.35	–8.82 to –3.87	<.001

[a]AF: Agoraphobia Free app.

[b]SF: Stress Free app.

Completers' Analysis

Data on completion were available from participants who consistently replied to the surveys and questionnaires (n=74). There were 56 participants who completed 80% or more of the assigned intervention; 25 of those received Agoraphobia Free and 31 received Stress Free. Of those who were deemed noncompleters, 11 were in the Agoraphobia Free arm and 7 in the Stress Free arm. There was no relationship between treatment arm and completion of intervention (χ^2_1=1.5, *P*=.22).

We examined differences in completers' symptom severity between the 2 intervention groups. We produced a linear mixed model, with random effect of participant, and fixed effects of time (baseline, midpoint, end point), group (Agoraphobia Free and Stress Free), and the interaction between time and group. In line with the intention-to-treat analysis, there were no significant differences between the 2 groups at end point or midpoint, as Table 5 shows. The within-group changes at each follow-up time point compared with baseline were significant in both groups (all *P*<.001). Therefore, there were significant reductions in symptom severity in both groups over time, but those reductions were equivalent across the 2 groups.

Clinical Significance

Wichmann et al [39] recently showed that an overall postintervention (ie, face-to-face CBT) decrease in the total PAS score of about 4 to 5 points represents a clear clinical change in terms of quality of life and functioning. This is slightly smaller than the change we observed in our study (6 points) for end point PAS scores compared with baseline. The decrease in symptom severity was even more marked in the completers' sample (7 points in the Stress Free group and 10 points in the Agoraphobia Free group). Therefore, it seems that both apps were overall successful in achieving clinically significant change.

Of the 74 participants who provided data at end point, 46 (62%) had a reduction of 5 or more points on the PAS at end point. A total of 25 (66%) participants in the Stress Free group and 21 (58%) participants in the Agoraphobia Free group improved at least 5 points on the PAS at end point.

Safety and Use of Apps

A total of 7 participants (Agoraphobia Free: n=3, Stress Free: n=4) commented that certain app components were mildly stressful (eg, the distraction games, background music). There were no reported adverse events experienced as a result of either intervention. Also, 7 participants commented that the Agoraphobia Free app was confusing to follow at certain points. Another participant explained that they could not use the treatment app as much as they wanted to because they had depression and did not feel motivated.

Table 5. Completers' analysis at end point (12 weeks) and midpoint (6 weeks), n=56.

Time point of PAS[a] score	Agoraphobia Free mean (SD)	Stress Free mean (SD)	Effect estimate		
			Difference	95% CI	*P* value
End point	21.58 (13.10)	23.73 (12.12)	–2.15	–6.21 to 1.91	.30
Midpoint	26.09 (11.15)	26.93 (10.40)	–0.84	–4.05 to 2.37	.61

[a]PAS: Panic and Agoraphobia Scale. The mean baseline PAS score for both groups was 31.23.

Discussion

Principal Findings

The results from this RCT showed that participants who received Agoraphobia Free did not improve more than those who received the Stress Free app. Both groups showed reductions in symptom severity over time that were statistically significant, but those reductions seemed to be equivalent across the 2 groups. Both treatment apps were safe to use and yielded similar completion rates. Moreover, completers of either intervention showed marked improvements in symptom severity, with a 10-point drop in the PAS at end point compared with baseline in the Agoraphobia Free group. However, findings from this analysis should be considered with caution, as we performed the analysis on a specific subgroup of participants, and there is the possibility that factors other than the intervention influenced the outcome (eg, participant expectations). In addition, the power of this analysis was very low (below 50%) due to the small sample size.

Throughout the trial, participant attrition was particularly high, as almost half of the participants recruited dropped out of the trial or stopped responding to emails. Dropout rates were similar to those reported in other Web-based trials of self-guided interventions [40,41]. Many participants dropped out after providing consent, which we had not expected. We had to exclude those participants from the analysis, as no baseline data were available. It is unlikely that this exclusion of participants could have led to bias, as participants at that point were unaware of the groups they were randomly allocated to, and thus attrition was random between the 2 groups. Similarly, participants who dropped out or stopped responding after receiving the intervention did not seem to differ from those who engaged in the trial, and attrition was equivalent across the 2 groups. The analytic strategy we chose is robust and uses all available data from each participant, producing less-biased results than other methods of analysis and data imputation [38]. Overall, the fact that participant attrition appeared random, in addition to the fact that we used data from all participants in the analysis, offer support and confidence in the validity of the study findings.

Limitations

The trial had a few limitations. There was no follow-up after the completion of the intervention. Therefore, it was not possible to investigate whether the improvements in symptom severity were maintained over time or whether there were any differences between the 2 groups after a few months of using the apps. The information collected regarding the characteristics of the sample was also limited, as there were no data on comorbid disorders, other psychological or physiological treatments, or other demographic characteristics such as ethnicity and computer literacy. This information might have provided better insights into the sample and could have been related to intervention efficacy. There was also limited information on app completion, as the information available relied on participant report and we were not permitted to extract app use data for participants individually. Moreover, participants could not be blinded, which is a common limitation in eHealth trials.

Another limitation of the study was the absence of a waitlist control group. The comparison of 2 active groups that shared very similar features tested the effects of agoraphobia-specific therapeutic elements over and above those of a generic anxiety-related treatment. There was no control, however, for the effect of time, and it is therefore not possible to conclude whether the improvements observed were not because of natural recovery processes or factors other than the intervention. Although a waitlist control group would be a necessary addition in a future trial in order to clearly establish treatment efficacy, the primary focus of this study was to demonstrate whether a disorder-specific mobile-based intervention is warranted in the treatment of agoraphobic symptoms, compared with a more generic approach addressing anxiety.

Despite its limitations, the study also presented certain strengths. First, it is, to our knowledge, the first RCT to directly compare an agoraphobia-specific app with a generic anxiety-related app. Second, the sample recruited online was characterized by severe clinical symptoms and especially high levels of agoraphobia, which shows that recruitment based on self-identification is feasible and reliable. Such a sample may have also been hard to reach in a traditional multicenter trial with assessors conducting screening interviews, which would have been time consuming and would require a lot of resources. The trial was minimally intrusive for the participants, as it requested a limited amount of information, and the tools used were short and easy to complete. This helped achieve fair recruitment and response rates without any explicit individual guidance. Third, an important strength of the study was its external validity. The apps were offered as they would be offered in real-world settings, without requiring any screening of users or close supervision. There were very few restrictions on who could participate, therefore making our findings more easily generalizable to an Internet population who would simply download the app if they thought it would be relevant to them. Fourth, the results of the study provide an insight into what we would realistically expect from an unguided mobile intervention, such as a substantial dropout rate, small symptom change, and highest efficacy for those who consistently engage with and complete the intervention.

Comparison With Prior Work

The findings of this study relate to previous studies on the treatment of anxiety disorders. The findings add to the evidence base of computerized cognitive behavioral interventions examined in anxiety disorders [21-23]. Our findings are also consistent with previous studies showing that completers of Web-based interventions for panic (with or without agoraphobia) benefit the most [40,42]. Importantly, the lack of a significant difference between the 2 groups indicates that a generic anxiety-related app and a diagnosis-specific app are equally effective in treating agoraphobic symptoms. This finding is in line with evidence supporting the efficacy of transdiagnostic (ie, unified, nonspecific) cognitive behavioral treatments in anxiety disorders. Transdiagnostic treatments are based on the premise that "commonalities across disorders outweigh the differences" [43]. A recent systematic review and meta-analysis showed that transdiagnostic treatments can be as efficacious as diagnosis-specific treatments in reducing anxiety symptoms and more effective than waitlist or attention control conditions [44]. Transdiagnostic CBT programs tested in different anxiety disorders (including panic disorder) can also be successfully administered via the Internet [45], producing effects equivalent to those of disorder-specific Internet-based interventions [46,47]. Moreover, transdiagnostic CBT does not differ from diagnosis-specific CBT in terms of treatment credibility [48]. Our study adds to this body of evidence suggesting that, for people with agoraphobia, an agoraphobia-specific app does not produce any additional benefits in relation to a transdiagnostic anxiety-targeting app.

Clinical Implications

The findings of this study have implications for clinical practice. This trial shows that mobile apps can be successfully administered to a particular population that is hard to reach otherwise, without requiring guidance by a clinician. This could potentially save time for clinicians, while it could also be more convenient for patients and help them overcome many barriers to treatment, such as traveling to sites or long waiting lists. Furthermore, a self-guided app could be easily introduced into a stepped care model as a minimal intervention, as it is less intense than a clinician-guided intervention or individual therapy [49]. Since there was no evidence of superiority of one type of intervention over the other in this study, a choice between a generic and a targeted approach could be ultimately based on clinical judgment or patient preference. However, as suggested by Norton and Barrera [48], a generic anxiety-targeting app might be more resource efficient and easier to implement and disseminate than a diagnosis-specific app.

Conclusions

Overall, this RCT suggests that transdiagnostic anxiety-targeting mobile apps can be as effective as disorder-specific apps for people with agoraphobic symptoms. Results show that individuals who identify as having agoraphobia do not benefit more from an agoraphobia-specific than from a transdiagnostic app. Future research conducted by independent research teams should replicate the results of this study, further investigating the possibility that mobile-based transdiagnostic interventions for anxiety can be as effective as current gold standard disorder-specific interventions. A trial with a larger sample size and a waitlist control group is warranted to establish intervention efficacy and cost effectiveness in this population. Additionally, the collection of more-extensive demographic and clinical information can help examine under which circumstances a diagnosis-specific or a generic approach is more appropriate. For example, future research could investigate whether patients with more severe symptoms or with comorbid disorders (eg, depression) benefit more from one type of intervention than the other. While there is still much to learn about treatment approaches in anxiety disorders, many studies, including this one, show that e-mental health interventions can overcome barriers and be effective in reducing clinical symptoms.

Acknowledgments

The authors would like to thank Dr Rebecca Jones, Senior Statistician at University College London, who provided statistical advice; Ana Cavero for her help and support of the project; and Kieron Kirkland and Kathy Marcham from Nominet Trust (funding body) for their support during the development of Agoraphobia Free. The authors would also like to thank all the participants who took part in the study.

Conflicts of Interest

JASF is the Chief Executive Officer of Thrive Therapeutic Software (trading name of Virtually Free Ltd), the company that developed and commercializes both of the apps used in this trial.

Multimedia Appendix 1

Agoraphobia Free session outline.

[PDF File (Adobe PDF File), 30KB - jmir_v19i11e398_app1.pdf]

Multimedia Appendix 2

Screenshots of Agoraphobia Free and Stress Free.

[PDF File (Adobe PDF File), 546KB - jmir_v19i11e398_app2.pdf]

Multimedia Appendix 3

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 538KB - jmir_v19i11e398_app3.pdf]

References

1. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
2. Alonso J, Lépine J, ESEMeD/MHEDEA 2000 Scientific Committee. Overview of key data from the European Study of the Epidemiology of Mental Disorders (ESEMeD). J Clin Psychiatry 2007;68 Suppl 2:3-9 [FREE Full text] [Medline: 17288501]
3. Emmelkamp PMG, Wittchen HU. Specific phobias. In: Andrews G, Charney CS, Sirovatka PJ, Regier DA, editors. Stressed-Induced and Fear Circuitry Disorders: Refining the Research Agenda for DSM-V. Arlington, VA: APA; 2009.
4. Skapinakis P, Lewis G, Davies S, Brugha T, Prince M, Singleton N. Panic disorder and subthreshold panic in the UK general population: epidemiology, comorbidity and functional limitation. Eur Psychiatry 2011 Sep;26(6):354-362. [doi: 10.1016/j.eurpsy.2010.06.004] [Medline: 20813508]
5. Kessler RC, Chiu WT, Jin R, Ruscio AM, Shear K, Walters EE. The epidemiology of panic attacks, panic disorder, and agoraphobia in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2006 Apr;63(4):415-424 [FREE Full text] [doi: 10.1001/archpsyc.63.4.415] [Medline: 16585471]
6. National Institute for Health and Care Excellence. Generalised anxiety disorder and panic disorder (with or without agoraphobia) in adults: management in primary, secondary and community care. Clinical guideline no. CG113. London, UK: National Institute for Health and Care Excellence; 2011 Jan 26. URL: https://www.nice.org.uk/guidance/cg113/resources/generalised-anxiety-disorder-and-panic-disorder-in-adults-management-pdf-35109387756997 [accessed 2017-11-14] [WebCite Cache ID 6uyEhbqZK]
7. Clark DM. A cognitive approach to panic. Behav Res Ther 1986;24(4):461-470. [doi: 10.1016/0005-7967(86)90011-2]
8. Bennett-Levy J. Oxford Guide to Behavioural Experiments in Cognitive Therapy. Oxford, UK: Oxford University Press; 2004.
9. Salkovskis PM. The importance of behaviour in the maintenance of anxiety and panic: a cognitive account. Behav Psychother 1991 Jun 16;19(01):6. [doi: 10.1017/S0141347300011472]
10. Telch M, Lancaster C. Is there room for safety behaviours in exposure therapy for anxiety disorders? In: Neudeck P, Wittchen HU, editors. Exposure Therapy: Rethinking the Model: Refining the Method. New York, NY: Springer; 2012:313-334.
11. Furukawa TA, Watanabe N, Churchill R. Psychotherapy plus antidepressant for panic disorder with or without agoraphobia: systematic review. Br J Psychiatry 2006 Apr;188:305-312 [FREE Full text] [doi: 10.1192/bjp.188.4.305] [Medline: 16582055]
12. Allen LB, White KS, Barlow DH, Shear MK, Gorman JM, Woods SW. Cognitive-behavior therapy (CBT) for panic disorder: relationship of anxiety and depression comorbidity with treatment outcome. J Psychopathol Behav Assess 2010 Jun;32(2):185-192 [FREE Full text] [doi: 10.1007/s10862-009-9151-3] [Medline: 20421906]
13. Marchand A, Roberge P, Primiano S, Germain V. A randomized, controlled clinical trial of standard, group and brief cognitive-behavioral therapy for panic disorder with agoraphobia: a two-year follow-up. J Anxiety Disord 2009 Dec;23(8):1139-1147. [doi: 10.1016/j.janxdis.2009.07.019] [Medline: 19709851]
14. Roberge P, Marchand A, Reinharz D, Savard P. Cognitive-behavioral treatment for panic disorder with agoraphobia: a randomized, controlled trial and cost-effectiveness analysis. Behav Modif 2008 May;32(3):333-351. [doi: 10.1177/0145445507309025] [Medline: 18391051]
15. Vos SPF, Huibers MJH, Diels L, Arntz A. A randomized clinical trial of cognitive behavioral therapy and interpersonal psychotherapy for panic disorder with agoraphobia. Psychol Med 2012 Dec;42(12):2661-2672. [doi: 10.1017/S0033291712000876] [Medline: 22717019]
16. Brown TA, Antony MM, Barlow DH. Diagnostic comorbidity in panic disorder: effect on treatment outcome and course of comorbid diagnoses following treatment. J Consult Clin Psychol 1995 Jun;63(3):408-418. [Medline: 7608353]
17. Andrews G, Issakidis C, Carter G. Shortfall in mental health service utilisation. Br J Psychiatry 2001 Nov;179:417-425 [FREE Full text] [Medline: 11689399]
18. Coles ME, Coleman SL. Barriers to treatment seeking for anxiety disorders: initial data on the role of mental health literacy. Depress Anxiety 2010;27(1):63-71. [doi: 10.1002/da.20620] [Medline: 19960488]
19. Alvidrez J, Azocar F. Distressed women's clinic patients: preferences for mental health treatments and perceived obstacles. Gen Hosp Psychiatry 1999;21(5):340-347. [Medline: 10572775]
20. Young AS, Klap R, Sherbourne CD, Wells KB. The quality of care for depressive and anxiety disorders in the United States. Arch Gen Psychiatry 2001 Jan 01;58(1):55. [doi: 10.1001/archpsyc.58.1.55]

21. Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. Health Technol Assess 2006 Sep;10(33):iii, xi-xiv, 1-iii, xi-xiv168 [FREE Full text] [Medline: 16959169]
22. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
23. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015;3:CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
24. Morris ME, Kathawala Q, Leen TK, Gorenstein EE, Guilak F, Labhard M, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010;12(2):e10 [FREE Full text] [doi: 10.2196/jmir.1371] [Medline: 20439251]
25. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
26. Rizvi SL, Dimeff LA, Skutch J, Carroll D, Linehan MM. A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. Behav Ther 2011 Dec;42(4):589-600. [doi: 10.1016/j.beth.2011.01.003] [Medline: 22035988]
27. Bandelow B. Assessing the efficacy of treatments for panic disorder and agoraphobia. II. The Panic and Agoraphobia Scale. Int Clin Psychopharmacol 1995 Jun;10(2):73-81. [Medline: 7673659]
28. Vøllestad J, Nielsen MB, Nielsen GH. Mindfulness- and acceptance-based interventions for anxiety disorders: a systematic review and meta-analysis. Br J Clin Psychol 2012 Sep;51(3):239-260. [doi: 10.1111/j.2044-8260.2011.02024.x] [Medline: 22803933]
29. Conrad A, Roth WT. Muscle relaxation therapy for anxiety disorders: it works but how? J Anxiety Disord 2007;21(3):243-264. [doi: 10.1016/j.janxdis.2006.08.001] [Medline: 16949248]
30. Bandelow B, Hajak G, Holzrichter S, Kunert HJ, Rüther E. Assessing the efficacy of treatments for panic disorder and agoraphobia. I. Methodological problems. Int Clin Psychopharmacol 1995 Jun;10(2):83-93. [Medline: 7673660]
31. Bandelow B, Brunner E, Broocks A, Beinroth D, Hajak G, Pralle L, et al. The use of the Panic and Agoraphobia Scale in a clinical trial. Psychiatry Res 1998 Jan 16;77(1):43-49. [Medline: 10710174]
32. Bandelow B, Broocks A, Pekrun G, George A, Meyer T, Pralle L, et al. The use of the Panic and Agoraphobia Scale (P & A) in a controlled clinical trial. Pharmacopsychiatry 2000 Sep;33(5):174-181. [doi: 10.1055/s-2000-12982] [Medline: 11071019]
33. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
34. Haahr M. Random.org. Dublin, Ireland: Randomness and Integrity Services Ltd; 2017. URL: https://www.random.org/ [accessed 2017-10-23] [WebCite Cache ID 6uR5mlyEi]
35. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
36. Legislation.gov.uk. Data Protection Act 1998 c. 29, Part I, Section 2. 1998. URL: http://www.legislation.gov.uk/ukpga/1998/29/section/2 [accessed 2017-11-16] [WebCite Cache ID 6v1JODMtC]
37. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
38. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
39. Wichmann S, Kirschbaum C, Lorenz T, Petrowski K. Effects of the cortisol stress response on the psychotherapy outcome of panic disorder patients. Psychoneuroendocrinology 2017 Mar;77:9-17. [doi: 10.1016/j.psyneuen.2016.11.030] [Medline: 27987430]
40. Van Ballegooijen BW, Riper H, Klein B, Ebert DD, Kramer J, Meulenbeek P, et al. An Internet-based guided self-help intervention for panic symptoms: randomized controlled trial. J Med Internet Res 2013;15(7):e154 [FREE Full text] [doi: 10.2196/jmir.2362] [Medline: 23896222]
41. Marks IM, Kenwright M, McDonough M, Whittaker M, Mataix-Cols D. Saving clinicians' time by delegating routine aspects of therapy to a computer: a randomized controlled trial in phobia/panic disorder. Psychol Med 2004 Jan;34(1):9-17. [Medline: 14971623]
42. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]

43. McEvoy P, Nathan P, Norton P. Efficacy of transdiagnostic treatments: a review of published outcome studies and future research directions. J Cogn Psychother 2009 Feb 01;23(1):20-33. [doi: 10.1891/0889-8391.23.1.20]
44. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
45. Titov N, Andrews G, Johnston L, Robinson E, Spence J. Transdiagnostic Internet treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2010 Sep;48(9):890-899. [doi: 10.1016/j.brat.2010.05.014] [Medline: 20561606]
46. Fogliati VJ, Dear BF, Staples LG, Terides MD, Sheehan J, Johnston L, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided internet-delivered treatment for panic disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2016 Apr;39:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2016.03.005] [Medline: 27003376]
47. Dear BF, Staples LG, Terides MD, Fogliati VJ, Sheehan J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for social anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2016 Aug;42:30-44 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.004] [Medline: 27261562]
48. Norton PJ, Barrera TL. Transdiagnostic versus diagnosis-specific CBT for anxiety disorders: a preliminary randomized controlled noninferiority trial. Depress Anxiety 2012 Oct;29(10):874-882 [FREE Full text] [doi: 10.1002/da.21974] [Medline: 22767410]
49. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17 [FREE Full text] [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]

Abbreviations

CBT: cognitive behavioral therapy
CCBT: computerized cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
HeLP: Health eLiving Partnership Ltd
PAS: Panic and Agoraphobia Scale
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 23.03.17; peer-reviewed by J Newby, K Kaipainen; comments to author 15.06.17; revised version received 09.09.17; accepted 27.09.17; published 24.11.17.

Please cite as:
Christoforou M, Sáez Fonseca JA, Tsakanikos E
Two Novel Cognitive Behavioral Therapy–Based Mobile Apps for Agoraphobia: Randomized Controlled Trial
J Med Internet Res 2017;19(11):e398
URL: http://www.jmir.org/2017/11/e398/
doi:10.2196/jmir.7747
PMID:29175809

Original Paper

Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial

Akio Mantani[1], MD, PhD; Tadashi Kato[2], MD; Toshi A Furukawa[3], MD, PhD[‡]; Masaru Horikoshi[4], PhD; Hissei Imai[3], MD, PhD; Takahiro Hiroe[5], MD, PhD; Bun Chino[6], MD, PhD; Tadashi Funayama[7], MD, PhD; Naohiro Yonemoto[8], MPH; Qi Zhou[9], PhD; Nao Kawanishi[10], PhD

[1]Mantani Mental Clinic, Hiroshima, Japan
[2]Aratama Kokorono Clinic, Nagoya, Japan
[3]Department of Health Promotion and Human Behavior, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[4]National Center of Neurology and Psychiatry, Kodaira, Japan
[5]Waseda Clinic, Kani, Japan
[6]Ginza Taimei Clinic, Tokyo, Japan
[7]Funayama Mental Clinic, Nagoya, Japan
[8]Department of Biostatistics, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[9]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
[10]Advanced Telecommunications Research Institute International, Kyoto, Japan
[‡]the FLATT Investigators

Corresponding Author:
Toshi A Furukawa, MD, PhD
Department of Health Promotion and Human Behavior
School of Public Health, Graduate School of Medicine
Kyoto University
Yoshida Konoe-cho
Sakyo-ku
Kyoto, 6068501
Japan
Phone: 81 757538941
Fax: 81 757534643
Email: furukawa@kuhp.kyoto-u.ac.jp

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2018/8/e11702/

Abstract

Background: In the treatment of major depression, antidepressants are effective but not curative. Cognitive behavioral therapy (CBT) is also effective, alone or in combination with pharmacotherapy, but accessibility is a problem.

Objective: The aim is to evaluate the effectiveness of a smartphone CBT app as adjunctive therapy among patients with antidepressant-resistant major depression.

Methods: A multisite, assessor-masked, parallel-group randomized controlled trial was conducted in 20 psychiatric clinics and hospitals in Japan. Participants were eligible if they had a primary diagnosis of major depression and were antidepressant-refractory after taking one or more antidepressants at an adequate dosage for four or more weeks. After a 1-week run-in in which participants started the medication switch and had access to the welcome session of the app, patients were randomized to medication switch alone or to medication switch plus smartphone CBT app via the centralized Web system. The smartphone app, called Kokoro-app ("kokoro" means "mind" in Japanese), included sessions on self-monitoring, behavioral activation, and cognitive restructuring presented by cartoon characters. The primary outcome was depression severity as assessed by masked telephone assessors with

the Patient Health Questionnaire-9 (PHQ-9) at week 9. The secondary outcomes included the Beck Depression Inventory-II (BDI-II) and Frequency, Intensity, and Burden of Side Effects Ratings (FIBSER).

Results: In the total sample (N=164), 81 participants were allocated to the smartphone CBT in addition to medication change and 83 to medication change alone. In the former group, all but one participant (80/81, 99%) completed at least half, and 71 (88%) completed at least six of eight sessions. In the intention-to-treat analysis, patients allocated the CBT app scored 2.48 points (95% CI 1.23-3.72, *P*<.001; standardized mean difference 0.40) lower on PHQ-9 than the control at week 9. The former group also scored 4.1 points (95% CI 1.5-6.6, *P*=.002) lower on BDI-II and 0.76 points (95% CI –0.05 to 1.58, *P*=.07) lower on FIBSER. In the per-protocol sample (comfortable with the smartphone app, still symptomatic, and adherent to medication with mild or less side effects after run-in), the intervention group (n=60) scored 1.72 points (95% CI 0.25-3.18, *P*=.02) lower on PHQ-9, 3.2 points (95% CI –0.01 to 6.3, *P*=.05) lower on BDI-II, and 0.75 points (95% CI 0.03-1.47, *P*=.04) lower on FIBSER than the control (n=57). The treatment benefits were maintained up to week 17.

Conclusions: This is the first study to demonstrate the effectiveness of a smartphone CBT in the treatment of clinically diagnosed depression. Given the merits of the mobile mental health intervention, including accessibility, affordability, quality control, and effectiveness, it is clinically worthwhile to consider adjunctive use of a smartphone CBT app when treating patients with antidepressant-resistant depression. Research into its effectiveness in wider clinical contexts is warranted.

Trial Registration: Japanese Clinical Trials Registry UMIN CTR 000013693; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000015984 (Archived by WebCite at http://www.webcitation.org/6u6pxVwik)

(J Med Internet Res 2017;19(11):e373) doi:10.2196/jmir.8602

KEYWORDS

major depressive disorder; pharmacotherapy-resistant depression; mobile phone app; cognitive behavioral therapy; eHealth

Introduction

Major depression is highly prevalent, debilitating, and costly [1-3]. It is predicted to be the leading cause of premature mortality and disability in high-income countries by 2030, and the third-leading cause in low- and middle-income countries [4]. Resources for and access to care by those who suffer remain constrained in high-income countries, and severely so in low- to middle-income ones [5,6].

Although antidepressant pharmacotherapy represents the mainstay of treatment of major depression [7], after several weeks of treatment only 50% show reduction by half or more in their depression severity and only 30% return to a euthymic state [8]. No standard approach in the management of treatment-refractory depression exists. Guideline recommendations include increasing the dose, switching to another antidepressant, or augmenting treatment with another pharmacological or psychological therapy [9,10].

Cognitive behavioral therapy (CBT) has proved an effective treatment of major depression either alone [11] or in combination with pharmacotherapy [12,13]. A standard course of CBT, however, requires 10 to 20 sessions, each lasting 45 to 60 minutes, with an adequately qualified professional. Therefore, its availability is limited everywhere in the world [14,15]. Telephone or videoconference CBT eliminates the burden of in-person visits, but requires approximately equal therapist time and equal competence.

Guided self-help CBT requires minimal to no clinician time and has proved of comparable effectiveness to its face-to-face versions [16]. Information and communication technologies (ICT) for self-help CBT, including computerized stand-alone software and Internet-based webpages, have shown promising results in initial trials [17]. However, one recent pragmatic trial using two widely known online Web-based CBT packages failed to demonstrate added value over usual care in primary care likely because adherence to the program was very low [18].

The dominant modality of ICT is evolving rapidly and the smartphone is now transforming people's lives across the world. In comparison with stand-alone or Internet-connected computers, the smartphone enhances portability and immediacy, making CBT fully accessible and therefore promising new dimensions of guided self-help. Mobile health apps are currently proliferating in the electronic world, with more than 165,000 health apps available online [19-21]. However, few apps have demonstrated quality [22,23] and no randomized controlled trial (RCT) has yet proved the benefits of a smartphone app in comparison with a control condition in the treatment of clinically diagnosed major depressive disorder. Several trials have examined the use of smartphone to monitor symptoms [24-26] or smartphone apps applying the CBT principles, but only among participants recruited from the general population who had reported elevated levels of depressive symptoms on self-reports [27-33]. A few have examined participants with diagnosed major depression, but only against active controls. One pilot RCT compared the original Web-based CBT against its smartphone version in 35 participants with depression confirmed through telephone diagnostic interview [34]. Another group of researchers developed a smartphone app for behavioral activation and tested it against its face-to-face full version [35] or against a mindfulness-based program [36] among 93 and 81 patients, respectively, who were recruited through advertisements in mass media but whose diagnosis of major depression was confirmed through telephone interview. In these studies, results did not differ significantly between the intervention and control groups.

Whether smartphone-based CBT can have any demonstrable value in the treatment of clinically diagnosed major depression is a pressing issue for patients, clinicians, and policy makers

around the world [22,23]. We have developed and pilot-tested a smartphone app, called Kokoro-app (*kokoro* means "mind" in Japanese), that is based on a CBT manual with demonstrated effectiveness in remote telephone or group formats in several RCTs either alone [37,38] or in combination with antidepressants [39]. This study represents the first RCT to examine adjunctive smartphone-based CBT to medication change among patients with major depression unresponsive to prior antidepressant treatments.

Methods

Design

The study was a multisite, assessor-masked, parallel-group RCT with a 1:1 allocation ratio.

Participants

A detailed description of the study protocol has been reported elsewhere [40] and is attached as Multimedia Appendix 1. The study, which was approved by the ethics committees of the participating centers and registered in the Japanese clinical trials registry (UMIN CTR 000013693), took place in 20 psychiatric clinics and hospitals across Japan between September 2014 and October 2016. Multimedia Appendix 2 provides the complete list of trial sites and investigators.

Eligible participants (1) were aged between 25 and 59 years, (2) had a primary diagnosis of major depressive disorder without psychotic features according to the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) [41] as ascertained by using the Primary Care Evaluation of Mental Disorders procedure [42], (3) were antidepressant-resistant, defined as scoring 10 or more on the Beck Depression Inventory-II (BDI-II) [37,43] after taking one or more antidepressants at an adequate dosage for four or more weeks (stage I, II, or III according to the criteria by Thase and Rush [44]), (4) had not been prescribed escitalopram or sertraline, or received CBT or interpersonal therapy for the index episode.

The study psychiatrists introduced the trial to the potentially eligible patients from among the patients they were seeing and invited them to participate voluntarily. After full disclosure of the trial contents and procedures, all participants provided written informed consent. No advertisement through the media was used.

Randomization

On entry, all participants started switching their antidepressant, had the Kokoro-app installed onto their iPhone, and had access to the welcome session, which mainly aimed at, after a brief description of CBT, training the participants in the use of the smartphone and its speech recognition. (When the participants did not own an iPhone, we lent one to them for the duration of the trial.) After this 1-week run-in, they had a telephone interview with the Patient Health Questionnaire-9 (PHQ-9) and the Frequency, Intensity, and Burden of Side Effects Ratings (FIBSER) and were classified into one of two samples. The per-protocol sample comprised those who had not or only partially responded since week 0 (PHQ-9 scores ≥5 [45] at week 1), who tolerated and adhered to the new antidepressant, and had no problem using the smartphone and Kokoro-app. The auxiliary sample included all others. The total sample (the per-protocol sample plus the auxiliary sample) would answer the real-world question of the value of smartphone CBT among the patients for whom the clinicians would initially consider prescribing the smartphone CBT, including those who would not tolerate the new medication or who may have some difficulty using the smartphone app, whereas the per-protocol sample would answer the question of the value of smartphone CBT under the narrower circumstances (ie, among patients who were able to follow the protocol and for whom the app would be expected to demonstrate its full effects).

After stratification by group, the participants were randomized 1:1 to the combined antidepressant switch plus smartphone CBT arm (intervention arm) or the antidepressant switch alone arm (control arm) using an automated Web program implementing the method of minimization. Therefore, the randomization was concealed. Clinics, number of antidepressants previously prescribed for the index episode (≥3 vs <3), and a total score of the PHQ-9 at week 1 (≥10 vs <10) were used as minimization variables.

During the telephone interview for the PHQ-9, if participants reported suicidal ideation for more than half the days or nearly every day of the previous 2 weeks, the interviewer immediately notified the staff in the central office, who then notified the responsible psychiatrist.

Interventions

The intervention group received both the antidepressant switch and the smartphone CBT, whereas the control group received only the antidepressant switch during the intervention period up to week 9. Details of each intervention are described subsequently.

Antidepressant Switch

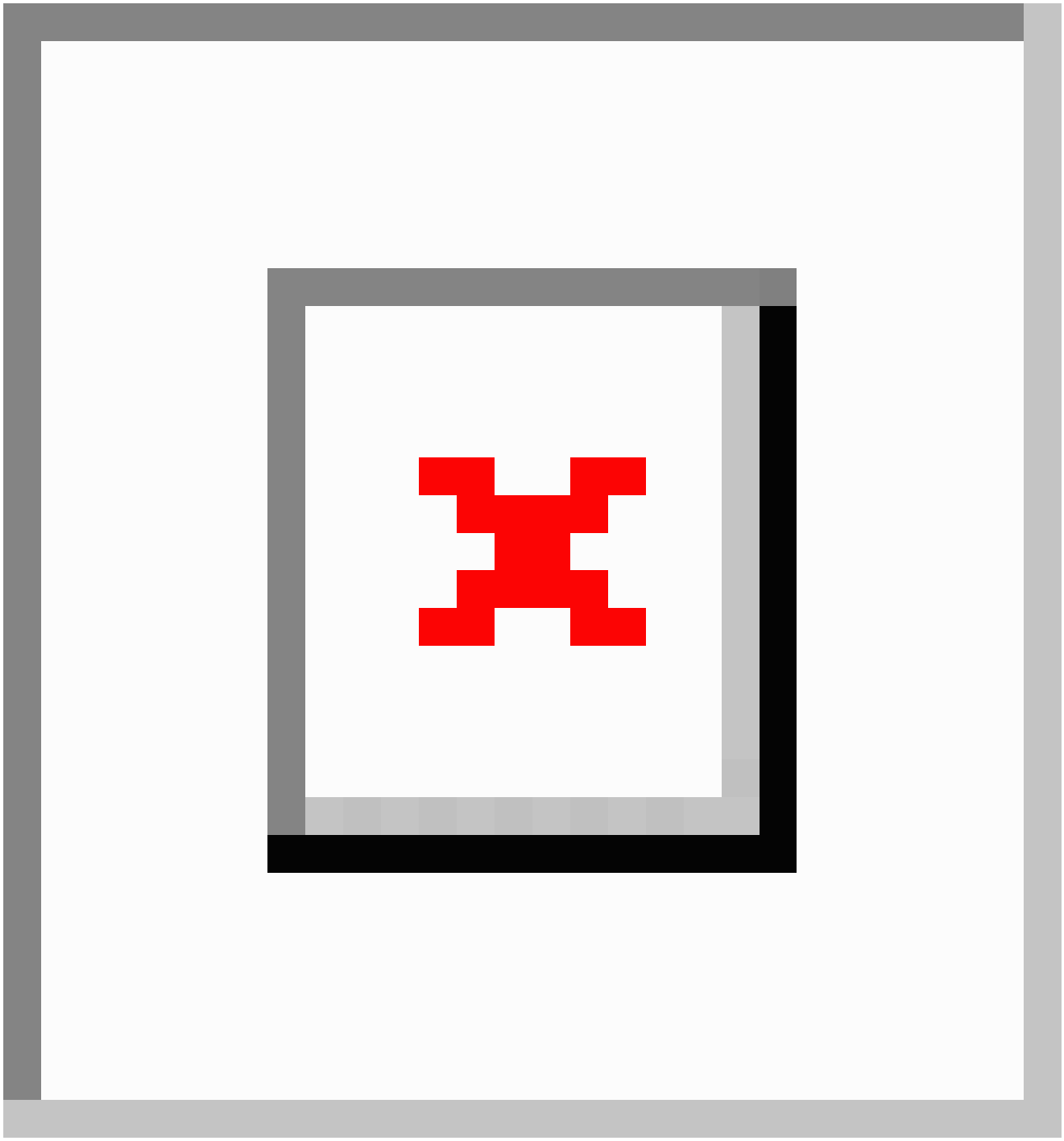
All study participants started switching their antidepressant either to escitalopram (5-10 mg/day) or to sertraline (25-100 mg/day) at week 0. The previous antidepressant was tapered off by week 5. We limited choice of antidepressants to escitalopram and sertraline, which showed a favorable profile in efficacy and acceptability in a previous systematic review [46], to ensure balance in antidepressant treatments during the trial. If the participants did not tolerate escitalopram or sertraline, the physician could revert to the previous antidepressant or start a new one. Only anxiolytics and hypnotics were allowed as coprescribed psychotropics. The frequency of visit was set at least once in four weeks, with additional visits as judged necessary by the study physician.

Smartphone Cognitive Behavioral Therapy

Kokoro-app is a self-help smartphone app consisting of eight sessions, including one welcome session, two sessions on self-monitoring, two sessions on behavioral activation, two sessions on cognitive restructuring, and an epilog focusing on relapse prevention. In each session, explanation of the principles and skills of CBT is provided in the format of instant messenger exchanges among cartoon characters (Figure 1). In the self-monitoring sessions, patients learn how to monitor their

reactions to situations in terms of feelings, thoughts, body reactions, and behaviors and describe them in "mind maps." In the behavioral activation sessions, patients learn to engage in "personal experiments" of small pleasurable actions according to the principle "When your body moves, so does your mind."

Figure 1. Screenshots from Kokoro-app.



To help patients broaden their thoughts, the cognitive restructuring sessions provide four tools, each of which guides the patients to alternative thoughts through interaction with the characters. Each session is supposed to take 1 week: participants can proceed to the next session only after 1 week and only after they have completed one homework from that session. See Multimedia Appendix 3 for more details. One session requires approximately 20 minutes to complete.

The progress of each participant on Kokoro-app could be monitored at "Kokoro-Web." Participants and their treating psychiatrists could access their records using a unique identification number and password. Secure Sockets Layer certified security of the data exchanged through the Internet.

Every week, the central office sent an email to participants to congratulate them on their progress. The message was templated but individually modified based on the progress of participants and the comments they uploaded at the end of each session. Writing one such email took 3 to 10 minutes.

All study participants had access to the welcome session at week 0. After randomization at week 1, the intervention group

received another password and continued with further sessions. The control group continued with the medication change only. Face-to-face CBT or interpersonal therapy were prohibited for either group.

Follow-Up Period After 9 Weeks

After assessments at week 9, there were no further restrictions in medications or psychotherapies. The participants in the control arm received a password to begin the Kokoro-app; follow-up assessments took place at week 17.

Outcomes

The masked assessors interviewed the participants by telephone with PHQ-9 and FIBSER at weeks 0, 1, 5, 9, and 17.

The primary outcome was the PHQ-9 at week 9. It consists of the nine diagnostic criteria items of major depression [47]. Each item is rated between 0="not at all" through 3="nearly every day." The total score ranges between 0 and 27. The instrument has excellent reliability, validity, and responsiveness [48]. Remission was defined as scoring four or less on PHQ-9, and response as 50% or greater reduction from baseline.

Secondary outcomes included the FIBSER, which assesses the frequency, intensity, and burden of side effects, each on a 7-point scale, with a total score between 3 and 27, [49] and the BDI-II [43], a self-report measure of depression severity that asks about 21 symptoms of depression, each on a scale between 0 and 3, with the total score between 0 and 63.

The study participants and the psychiatrists in charge of medication change were aware of the treatment allocation. The outcome assessors conducting telephone interviews were unaware of allocation. The success of this masking was evaluated by calculating the Bang Index [50] of assessors' treatment guesses after each telephone assessment. The Bang Index is scaled to an interval of –1 to 1, 1 being complete lack of blinding, 0 being consistent with perfect random guessing, and –1 indicating opposite guessing.

Sample Size

The study was powered to detect a moderate effect size of 0.5 in terms of standardized mean difference between the two treatments for the primary outcome at week 9, with 80% power at two-sided alpha level of .05. Assuming that 30% would leave the study or be classified into the auxiliary group at week 1, the required total sample size was 164 participants.

Statistical Analysis

All analyses were undertaken according to the intention-to-treat principle, including all the randomized participants for the total sample as well as for the per-protocol sample.

For each continuous outcome up to week 9, we used a linear mixed model including sites and patients as random effects and time (5 and 9 weeks), treatment, and time*treatment interaction, adjusting for its baseline score and the stratification variables, as fixed effects. The primary endpoint was the estimate of the least squares mean difference along with the 95% confidence interval (95% CI) at week 9. For the continuous outcome at week 17 follow-up, we used the similar linear mixed model but without the time*treatment interaction because this was a one-time comparison after all the participants received the smartphone CBT both in the intervention and the control groups. For the dichotomous outcomes, we used a generalized linear mixed model with the same random effects and the fixed effects. We chose odds ratios as the measure of effect. We used SAS version 9.4 (SAS Institute Inc, Cary, NC, USA). Multimedia Appendix 1 provides the statistical analysis plan.

Blinded Interpretation of the Results

The statistician (QZ), blinded to allocation, conducted the statistical analyses. The writing committee reviewed a statistical report in which the two treatment arms were designated A and B, and developed interpretation of the results and associated conclusions under two different scenarios, one assuming A to be the smartphone CBT plus medication change arm and B to be the medication change alone arm, and another alternative scenario. The code was broken only after the writing committee signed off on the agreed-on interpretations (see Multimedia Appendix 4).

Results

Participant Characteristics

Figure 2 shows the flow of participants through the study. Between September 2, 2014, and July 1, 2016, 323 patients were assessed for eligibility; 166 patients provided informed consent and started medication change. Two withdrew consent before randomization at week 1. Therefore, we recruited 164 patients, of whom 117 found no difficulty with the smartphone, were adherent to the protocol treatment, remained at least moderately symptomatic, and constituted the per-protocol sample; 60 were allocated to the smartphone CBT and 57 to medication change alone. Of the remaining 47 participants in the auxiliary group, 21 were allocated to the intervention group and 26 to control. Primary outcome data at 9 weeks were obtained from all but one randomized participant (163/164, 99.4%).

Table 1 shows that the baseline demographic and clinical characteristics of the treatment groups were well balanced. Typically, patients were in their thirties to forties, had three previous episodes, were in the current episode for nearly 2 years, and were severely to moderately depressed.

Figure 2. Assessment, randomization, and follow-up of study participants. CBT: cognitive behavioral therapy; PHQ-9: Patient Health Questionnaire 9; TAU: treatment as usual.

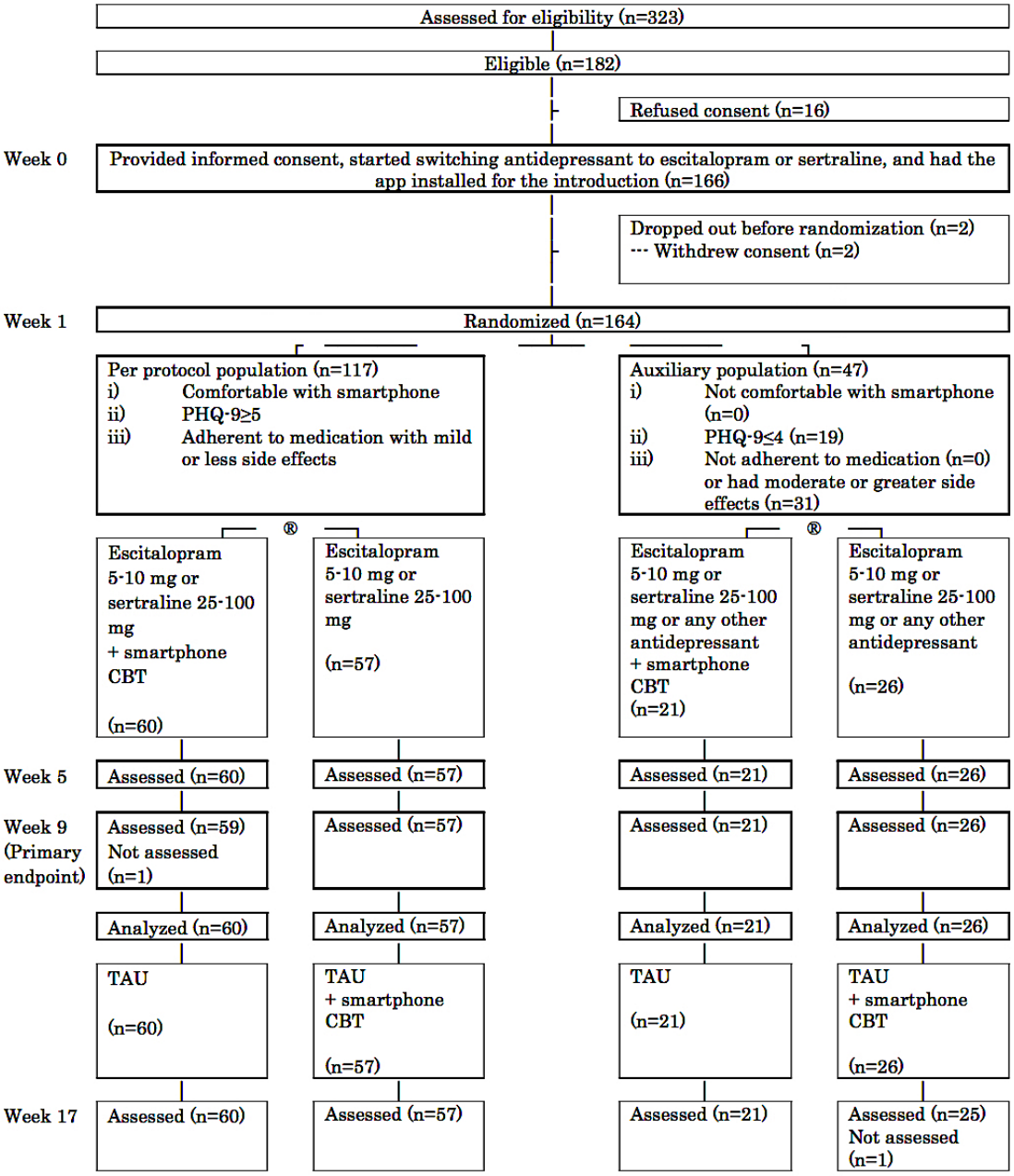


Table 1. Baseline characteristics of the intention-to-treat samples.

Characteristics	Total sample (N=164)		Per-protocol sample (n=117)	
	Smartphone CBT + medication change (n=81)	Medication change alone (n=83)	Smartphone CBT + medication change (n=60)	Medication change alone (n=57)
Demographic characteristics				
Age (years), mean (SD)	40.2 (8.8)	41.6 (8.9)	40.1 (9.0)	41.2 (8.6)
Sex (female), n (%)	46 (57)	41 (50)	35 (58)	31 (54)
Education (years), mean (SD)	14.6 (2.5)	14.9 (2.7)	14.8 (2.5)	15.1 (2.9)
Employment status, n (%)				
Employed full-time	34 (42)	29 (35)	27 (45)	21 (37)
Employed part-mime	7 (9)	5 (6)	5 (8)	5 (9)
On medical leave	21 (26)	30 (36)	14 (23)	21 (37)
Housewife	6 (7)	5 (6)	5 (8)	3 (5)
Student	0	2 (21)	0	1 (2)
Retired	0	0	0	0
Not employed	13 (16)	12 (15)	9 (15)	6 (11)
Marital status, n (%)				
Single, never married	34 (42)	31 (38)	24 (40)	22 (39)
Single, divorced, separated or widowed	13 (16)	7 (8)	12 (20)	5 (9)
Married	34 (42)	45 (54)	24 (40)	30 (53)
Baseline clinical characteristics, mean (SD)				
Age of onset at first episode (years)	31.8 (10.8)	34.6 (10.0)	32.2 (11.0)	34.0 (10.2)
Number of previous depressive episodes	3.4 (4.9)	3.0 (4.5)	3.5 (4.7)	2.8 (4.6)
Length of current episode (months)	24.2 (46.3)	23.0 (46.5)	27.0 (52.6)	22.3 (42.5)
PHQ-9 [a]				
Week 0	13.5 (5.5)	12.9 (5.3)	14.0 (5.2)	13.8 (5.1)
Week 1	12.6 (6.2)	11.9 (5.9)	13.4 (5.6)	12.6 (5.5)
BDI-II [a]				
Week 0	28.2 (11.2)	26.2 (11.0)	29.4 (10.6)	27.4 (10.7)
Week 1	26.2 (11.7)	24.7 (12.2)	28.1 (11.0)	26.2 (11.6)
FIBSER [a]				
Week 0	4.8 (4.5)	5.2 (3.1)	4.4 (2.8)	5.4 (2.9)
Week 1	6.4 (4.5)	6.8 (4.4)	5.0 (3.0)	5.2 (2.5)

[a]BDI-II: Beck Depression Inventory-II; FIBSER: Frequency, Intensity, and Burden of Side Effects Ratings; PHQ-9: Patient Health Questionnaire 9.

Treatments Received

In the total sample, 16 of 81 participants (20%) stopped the protocol antidepressant treatment by escitalopram or sertraline in the intervention arm, as did 14 of 83 (17%) in the medication only arm (Table 2). Some received augmentation drugs, such as antipsychotics or lithium, which were not allowed in the protocol for the antidepressant switch (n=15 in the total sample), a few had deterioration or side effects and could not continue the protocol treatment (n=5), and a few others got so well and did not want to continue with the protocol medication (n=2). The medication dosages were comparable between the arms.

For the smartphone CBT, all but one participant (80/81, 99%) completed at least half the sessions of the program, and 71 of 81 (88%) completed at least six of eight sessions. It took the patients, on average, 10.8 (SD 4.2) days to complete one session. The patients filled in a mean 11.2 (SD 11.4) “mind maps” for self-monitoring, conducted a mean 14.4 (SD 17.1) behavioral activation “personal experiments,” and generated a mean 6.1 (SD 6.0) alternative thoughts for cognitive restructuring (Table 2).

Table 2. Treatment received in the intention-to-treat samples.

Therapy	Total sample (N=164)		Per-protocol sample (n=117)	
	Smartphone CBT + medication change (n=81)	Medication change alone (n=83)	Smartphone CBT + medication change (n=60)	Medication change alone (n=57)
Pharmacotherapy				
Discontinuation of protocol antidepressant treatment by escitalopram or sertraline by week 9, n (%)	16 (20)	14 (17)	10 (17)	8 (14)
Discontinuation reason, n				
Prescription of prohibited drugs	6	9	4	5
Amelioration	1	1	—	1
Deterioration	1	—	1	—
Side effects	3	1	1	—
Other	5	3	4	2
Discontinuation of any antidepressant therapy by week 9, n (%)	5 (6)	2 (2)	2 (3)	2 (4)
Escitalopram dosage at week 9 (mg/day), mean (SD)	9.5 (3.0), n=48	10.0 (3.2), n=49	9.7 (3.1), n=38	10.1 (2.9), n=34
Sertraline dosage at week 9 (mg/day), mean (SD)	79.0 (26.7), n=25	83.6 (23.4), n=29	81.6 (26.1), n=19	83.0 (24.9), n=22
Smartphone CBT				
Sessions completed, n		—		—
0	1		—	
3	1		1	
4	4		3	
5	4		2	
6	11		8	
7	17		13	
8	43		33	
Time per session (days), mean (SD)	10.8 (4.2)	—	10.7 (4.0)	—
Number of mind maps for self-monitoring, mean (SD)	11.2 (11.4)	—	10.8 (10.3)	—
Number of behavioral activation tasks, mean (SD)	14.4 (17.1)	—	14.6 (17.8)	—
Number of alternative thoughts, mean (SD)	6.1 (6.0)	—	6.5 (6.0)	—

Table 3. Outcomes at weeks 9 and 17 for the total sample (N=164).

Outcomes[a]	Smartphone CBT + medication change (n=81), mean/% (95% CI)	Medication change alone (n=83), mean/% (95% CI)	Adjusted difference/OR (95% CI)[b]	*P* value
End of randomized trial (week 9)				
PHQ-9	7.94 (6.98, 8.89)	10.41 (9.45, 11.33)	–2.48 (–3.72, –1.23)	<.001
Remission	30.5% (19.7%, 43.9%)	17.8% (10.3%, 29.0%)	2.02 (0.93, 4.42)	.08
Response	42.3% (29.4%, 56.4%)	21.2% (12.7%, 33.2%)	2.73 (1.35, 5.53)	.005
BDI-II	19.3 (17.0, 21.5)	23.3 (21.6, 25.5)	–4.1 (–6.6, –1.5)	.002
FIBSER	4.38 (3.72, 5.03)	5.14 (4.52, 5.76)	–0.76 (–1.58, 0.05)	.07
Follow-up (week 17)				
PHQ-9	7.95 (6.73, 9.17)	8.76 (7.58, 9.95)	–0.81 (–2.24, 0.62)	.26
BDI-II	17.2 (14.4, 20.0)	19.1 (16.4, 21.8)	–1.9 (–4.9, 1.2)	.22
FIBSER	4.62 (3.83, 5.42)	5.10 (4.32, 5.89)	–0.48 (–1.34, 0.37)	.27

[a]BDI-II: Beck Depression Inventory-II, FIBSER: Frequency, Intensity, and Burden of Side Effects Ratings, PHQ-9: Patient Health Questionnaire 9.

[b]For each continuous outcome up to week 9, we used a linear mixed model including sites and patients as random effects and time (5 and 9 weeks), treatment, and time*treatment interaction, adjusting for its baseline score and the stratification variables, as fixed effects. For the continuous outcome at week 17 follow-up, we used the similar linear mixed model but without time*treatment interaction. For the dichotomous outcomes, the generalized linear mixed model was used along with the same random effects and the fixed effects. The summary effect measures are adjusted score differences for PHQ-9, BDI-II, and FIBSER, and are odds ratios for remission and response.

Outcomes for the Total Sample

In the intention-to-treat analysis of the total sample, patients who were allocated the CBT app (n=81) scored 2.48 points (95% CI 1.23-3.72, *P*<.001; standardized mean difference [SMD] 0.40) lower on PHQ-9 than those who were not (n=83) at week 9 (Table 3, Figure 3). The former group also scored 4.1 points (95% CI 1.5-6.6, *P*=.002) lower on BDI-II. Significantly more participants showed response (OR 2.73, 95% CI 1.35-5.53, *P*=.005). However, the increase in remission did not reach statistical significance (OR 2.02, 95% CI 0.93-4.42, *P*=.08).

With regard to the harm outcomes, patients using the smartphone CBT reported somewhat less overall burden of side effects, but the difference was not statistically significant (FIBSER mean difference=–0.76, 95% CI –1.58 to 0.05, *P*=.07). There was one report of suicidality (self-injurious behavior without suicidal intent) in the combined treatment arm and one report of a serious adverse event in the control arm (brief hospital admission for examination of preexisting spinal canal stenosis).

Outcomes for the Per-Protocol Sample

In the per-protocol sample, who were comfortable with the smartphone app, were still symptomatic, and were adherent to medication with mild or less side effects after the run-in, the patients who received smartphone CBT in addition to medication change (n=60) scored mean 1.72 (95% CI 0.25-3.18, *P*=.02; SMD 0.28) points lower on PHQ-9 than those undergoing medication change alone (n=57) at week 9. The combined treatment arm was superior to the control arm in terms of BDI-II (difference=–3.2, 95% CI –6.3 to 0.0, *P*=.05, not statistically significant), but not in terms of remission (OR 1.99, 95% CI 0.74-5.38, *P*=.17) or response (OR 2.11, 95% CI 0.92-4.85, *P*=.08).

In terms of the harm outcome, the combination treatment arm reported significantly less overall burden of side effects (FIBSER mean difference=–0.75, 95% CI –1.47 to –0.03, *P*=.04) (Table 4, Figure 3).

Follow-Up at Week 17

The participants who had access to the smartphone app in the first 9 weeks maintained their gains for a further 8 weeks. At week 17, when the participants in the control arm also had access to the smartphone app, the results were comparable between the two groups in terms of PHQ-9, BDI-II, and FIBSER, both for the total sample and for the per-protocol sample (Tables 3 and 4).

Masking

The Bang Index of the treatment guesses by the masked assessors at weeks 5, 9, and 17 was 0.10 (95% CI –0.13 to 0.33), 0.29 (95% CI 0.06-0.51), and 0.30 (95% CI 0.08-0.52) for the intervention arm and –0.21 (95% CI –0.45 to 0.03), –0.18 (95% CI –0.41 to 0.06), and –0.30 (95% CI –0.53 to –0.07) for the control arm, respectively. The observed patterns indicate that the raters guessed the treatment allocation haphazardly or guessed it to be the smartphone CBT arm more often regardless of the actual allocation, resulting in ideally unbiased assessment of outcomes in the trial [51].

Figure 3. Trajectory of the Patient Health Questionnaire-9 (PHQ-9), Beck Depression Inventory-II (BDI-II), and Frequency, Intensity, and Burden of Side Effects Ratings (FIBSER) scores for the intervention (blue line) and control (red line) groups in the total (N=164) and per-protocol (n=117) samples. Error bars show standard errors for model-based least squares means.

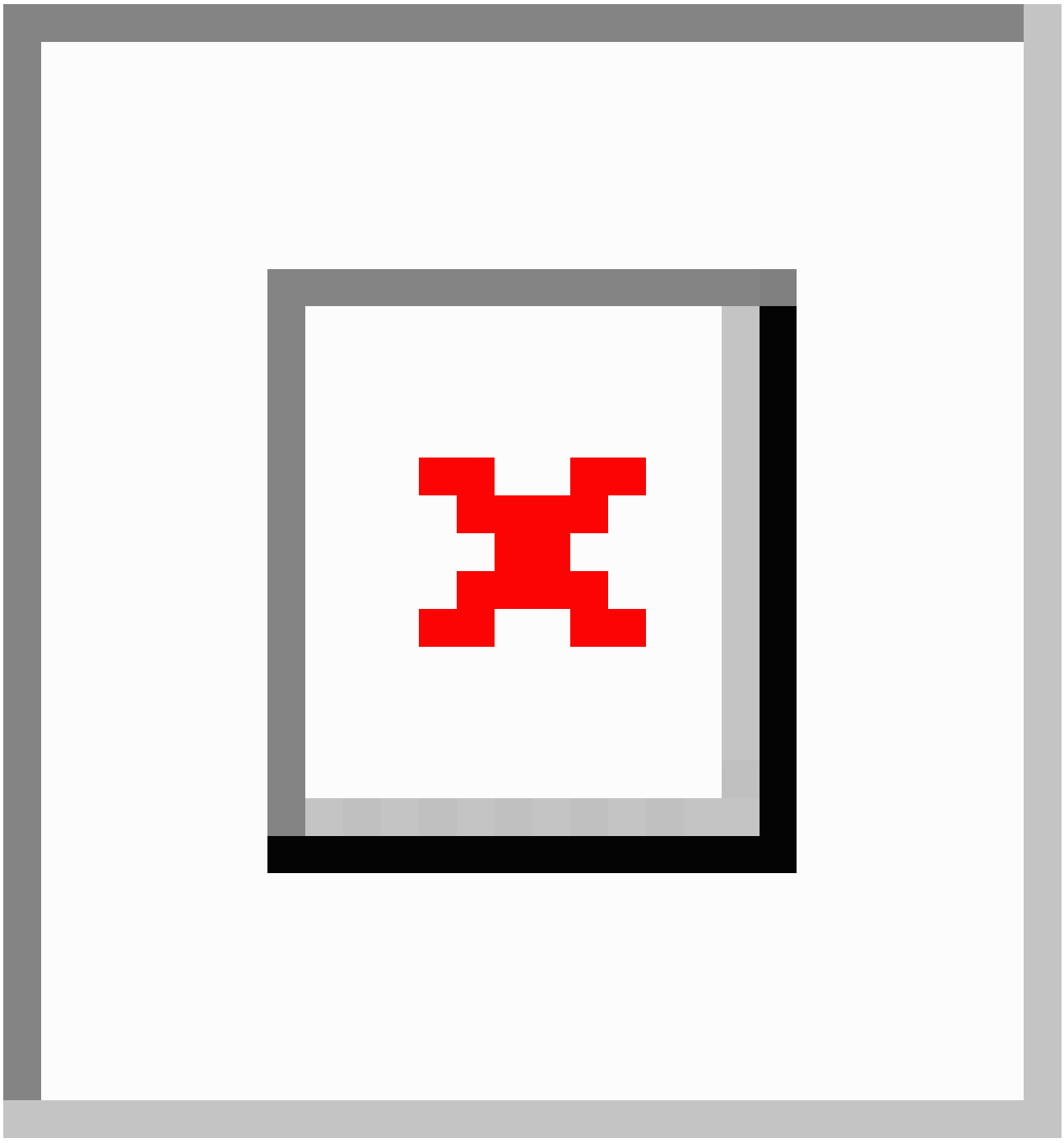


Table 4. Outcomes at weeks 9 and 17 for the per-protocol sample (n=117).

Outcomes[a]	Smartphone CBT + medication change (n=60), mean/% (95% CI)	Medication change alone (n=57), mean/% (95% CI)	Adjusted difference/OR (95% CI)[b]	*P* value
End of randomized trial (week 9)				
PHQ-9	8.92 (7.81, 10.03)	10.64 (9.52, 11.76)	–1.72 (–3.18, –0.25)[c]	.02
Remission	18.2% (8.5%, 34.8%)	10.0% (4.0%, 23.2%)	1.99 (0.74, 5.38)	.17
Response	31.6% (18.7%, 48.3%)	18.0% (9.2%, 32.3%)	2.11 (0.92, 4.85)	.08
BDI-II	21.0 (18.6, 23.5)	24.2 (21.7, 26.8)	–3.2 (–6.3, 0.0)	.05
FIBSER	4.11 (3.44, 4.78)	4.86 (4.18, 5.53)	–0.75 (–1.47, –0.03)	.04
Follow-up (week 17)				
PHQ-9	8.92 (7.40, 10.44)	8.85 (7.31, 10.39)	0.07 (–1.68, 1.82)	.94
BDI-II	19.4 (15.9, 22.9)	20.0 (16.5, 23.6)	–0.6 (–4.4, 3.1)	.75
FIBSER	4.14 (3.40, 4.88)	4.46 (3.71, 5.21)	–0.32 (–1.28, 0.63)	.50

[a]BDI-II: Beck Depression Inventory-II; FIBSER: Frequency, Intensity, and Burden of Side Effects Ratings; PHQ-9: Patient Health Questionnaire 9.

[b]For each continuous outcome up to week 9, we used a linear mixed model including sites and patients as random effects and time (5 and 9 weeks), treatment, and time*treatment interaction, adjusting for its baseline score and the stratification variables, as fixed effects. For the continuous outcome at week 17 follow-up, we used the similar linear mixed model but without time*treatment interaction. For the dichotomous outcomes, the generalized linear mixed model was used along with the same random effects and the fixed effects. The summary effect measures are adjusted score differences for PHQ-9, BDI-II, and FIBSER, and are odds ratios for remission and response.

[c]Primary endpoint per protocol.

Discussion

In patients with major depression who had not responded to one or more antidepressants, adding smartphone CBT to medication change was more effective than treatment by medication change alone. The smartphone CBT also decreased the overall side effect burden of the pharmacotherapy.

The magnitude of benefit of the adjunctive mobile CBT was approximately 2 points on the PHQ-9 and 3 to 4 points on the BDI-II. Using the observed standard deviation at week 9, these differences translate into standardized mean differences of 0.28 to 0.40 and are comparable to that of 0.31 for antidepressants over placebo reported in a comprehensive systematic review of phase II or III RCTs [52]. The remission and response rates almost doubled, corresponding with numbers needed to treat between 5 and 12.

Although several smartphone CBT apps have been experimented among general population participants with elevated self-reported symptoms [27-33], Kokoro-app is the first smartphone CBT app to prove to be effective in a RCT in comparison with an alternative treatment for patients suffering from clinically diagnosed major depression. Advantages of smartphone CBT include high accessibility, efficiency, and affordability. Further, it is less susceptible to quality control problems that may plague face-to-face therapies [53].

Unexpectedly, Kokoro-app also reduced the global burden of side effects due to pharmacotherapy. We speculate that the smartphone CBT, through which the patient actively searches for ways to overcome their depression, may increase their sense of self-control and decrease the subjective burden of side effects in comparison with standard pharmacotherapy.

All benefits were larger in the total sample than in the per-protocol comparison, where we had anticipated a larger effect at the protocol stage. Participants were not included in the per-protocol sample mainly because they suffered from side effects from the medication change (31 of 47, see Figure 1). The CBT app may be particularly beneficial to those patients who experience significant side effects with medication.

Strengths of this study include the concealed randomization, the successful masking of the outcome assessors, the close to 100% follow-up, the stratified randomization that allowed assessment of intervention impact in two key samples, and corroboration of the secondary outcomes including patients' self-reports. We followed the participants for 8 weeks after the end of the randomized comparison: the participants in the active intervention arm maintained the benefits, and the participants in the control arm who had access to Kokoro-app improved. Finally, the uptake of the CBT sessions via smartphone was satisfactory, with close to 90% of the participants finishing at least six of eight program sessions and actively engaging in homework tasks.

This study is not without limitations. First, it is possible that participants in the control condition, who were not allowed access to CBT sessions at the beginning but only after the waiting period, may have suffered "disappointment effect" through the intervention period. We tried to mitigate this limitation that may be common to many waiting list-controlled trials by not making our control condition a simple waiting list control in which the participants are not allowed to change their treatment [54], but an active medication change, which is one of the recommended treatment options for antidepressant-refractory patients. Nonetheless, the possibility of some contribution of disappointment effect among the control group cannot be negated. Secondly, it was impossible to mask

the participants and the clinicians administering drug treatment to the treatment allocation. However, we employed the remote telephone assessors to conduct assessments of the primary outcome, which resulted in successful masking. Lastly, it must be pointed out that in this refractory population, although the smartphone CBT in conjunction with medication switch substantively decreased depression severity, approximately 70% to 80% were still not remitted after 2 months of the combined treatment; additional face-to-face standard CBT may be helpful for these remaining patients.

These findings have demonstrated the effectiveness of smartphone CBT as an adjunctive intervention for antidepressant-resistant major depression. Given the merits of the mobile mental health intervention, including accessibility, affordability, quality control, and effectiveness, patients and clinicians may wish to use smartphone CBT as an adjunctive intervention when their depression does not respond adequately to antidepressant treatment alone. Further research of its effectiveness in wider clinical contexts, including its use as a stand-alone treatment in major depression and its role in relapse prevention, and in public health contexts such as its use for subthreshold depressive states or its utility in low- and middle-income countries is warranted. Elucidation of the effective components of the smartphone CBT package and their appropriate dosages and their integration within the existing health care systems also constitute areas needing further research.

Acknowledgments

The study was funded by the Ministry of Health, Labor and Welfare, Japan (H-22-Seishin-Ippan-008) from April 2013 through March 2014, and thereafter by the Japan Foundation for Neuroscience and Mental Health (JFNMH). The JFNMH received donations from Asahi Kasei, Eli Lilly, GSK, Janssen, MSD, Meiji Seika, Mochida, Otsuka, Pfizer, Shionogi, Taisho, and Mitsubishi-Tanabe. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

TAF and QZ 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. Study concept and design: TAF, MH, and NY. Acquisition, analysis, or interpretation of data: AM, TK, TAF, MH, HI, TH, BC, TF, NY, QZ, and NK. Drafting of the manuscript: TAF. Critical revision of the manuscript for important intellectual content: AM, TK, TAF, MH, HI, TH, BC, TF, NY, QZ, and NK. Statistical analysis: QZ and TAF. Obtained funding: TAF. Administrative, technical, or material support: HI, NY, and NK. Study supervision: TAF and MH.

Conflicts of Interest

AM has received lecture fees from Mochida, Eli Lilly, and Meiji. TK has received lecture fees from Eli Lilly and Mitsubishi-Tanabe, and has contracted research with GlaxoSmithKline, MSD, and Mitsubishi-Tanabe. TAF has received lecture fees from Eli Lilly, Janssen, Meiji, MSD, Otsuka, Pfizer, and Mitsubishi-Tanabe, and consultancy fees from Takeda Science Foundation. He has received research support from Mochida and Mitsubishi-Tanabe. TH has received lecture fees from Otsuka and MSD. BC has received lecture fees from Eli Lilly, Meiji, and Mitsubishi-Tanabe. TF has received lecture fees from Eli Lilly, Meiji, MSD, Otsuka, Pfizer, and Mitsubishi-Tanabe. TAF and MH developed the Kokoro-app. All the other authors declare that they have no competing interests.

Multimedia Appendix 1

Original study protocol (August 2014) and statistical analysis plan (July 2016).

[PDF File (Adobe PDF File), 275KB - jmir_v19i11e373_app1.pdf]

Multimedia Appendix 2

FLATT investigators and committee members.

[PDF File (Adobe PDF File), 77KB - jmir_v19i11e373_app2.pdf]

Multimedia Appendix 3

Contents of Kokoro-app.

[PDF File (Adobe PDF File), 824KB - jmir_v19i11e373_app3.pdf]

Multimedia Appendix 4

Blinded data analyses statement of interpretation.

[PDF File (Adobe PDF File), 128KB - jmir_v19i11e373_app4.pdf]

Multimedia Appendix 5

CONSORT‐EHEALTH checklist (V.1.6.1).

[PDF File (Adobe PDF File), 8MB - jmir_v19i11e373_app5.pdf]

References

1. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med 2011 Jul 26;9:90 [FREE Full text] [doi: 10.1186/1741-7015-9-90] [Medline: 21791035]
2. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]
3. Bloom D, Cafiero E, Jane-Llopis E, Abrahams-Gessel S, Bloom L, Fathima S. The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum; 2011.
4. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
5. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 2004 Jun 2;291(21):2581-2590. [doi: 10.1001/jama.291.21.2581] [Medline: 15173149]
6. Collins PY, Patel V, Joestl SS, March D, Insel TR, Daar AS, Scientific Advisory Board and the Executive Committee of the Grand Challenges on Global Mental Health, et al. Grand challenges in global mental health. Nature 2011 Jul 06;475(7354):27-30 [FREE Full text] [doi: 10.1038/475027a] [Medline: 21734685]
7. Marcus SC, Olfson M. National trends in the treatment for depression from 1998 to 2007. Arch Gen Psychiatry 2010 Dec;67(12):1265-1273. [doi: 10.1001/archgenpsychiatry.2010.151] [Medline: 21135326]
8. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 2006 Jan;163(1):28-40. [doi: 10.1176/appi.ajp.163.1.28] [Medline: 16390886]
9. American Psychiatric Association Work Group on Major Depressive Disorder. Practice Guideline for the Treatment of Patients with Major Depressive Disorder. Third Edition. Arlington, VA: American Psychiatric Association; 2010.
10. NICE. Depression: The Treatment and Management of Depression in Adults (partial update of NICE clinical guideline 23). London: National Institute for Clinical Excellence; 2009.
11. Furukawa T, Weitz E, Tanaka S, Hollon S, Hofmann S, Andersson G, et al. Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. Br J Psychiatry 2017 Mar;210(3):190-196. [doi: 10.1192/bjp.bp.116.187773] [Medline: 28104735]
12. Cuijpers P, Dekker J, Hollon SD, Andersson G. Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: a meta-analysis. J Clin Psychiatry 2009 Sep;70(9):1219-1229. [doi: 10.4088/JCP.09r05021] [Medline: 19818243]
13. Cuijpers P, van Straten A, Warmerdam L, Andersson G. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. Depress Anxiety 2009;26(3):279-288. [doi: 10.1002/da.20519] [Medline: 19031487]
14. Wang PS, Aguilar-Gaxiola S, Alonso J, Angermeyer MC, Borges G, Bromet EJ, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 2007 Sep 8;370(9590):841-850 [FREE Full text] [doi: 10.1016/S0140-6736(07)61414-7] [Medline: 17826169]
15. Olfson M, Marcus SC. National trends in outpatient psychotherapy. Am J Psychiatry 2010 Dec;167(12):1456-1463. [doi: 10.1176/appi.ajp.2010.10040570] [Medline: 20686187]
16. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
17. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided Internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
18. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015;351:h5627 [FREE Full text] [Medline: 26559241]
19. IMS Institute for Healthcare Informatics. 2015. Patient options expand as mobile healthcare apps address wellness and chronic disease treatment needs URL: http://www.imshealth.com/en/thought-leadership/quintilesims-institute/reports/patient-options-expand-as-mobile-healthcare-apps-address-wellness-and-chronic-disease-treatment-needs [accessed 2017-07-29] [WebCite Cache ID 6sJIcCKHV]

20. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
21. 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]
22. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 7;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
23. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 01;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
24. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
25. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
26. Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. BMC Fam Pract 2011;12:131 [FREE Full text] [doi: 10.1186/1471-2296-12-131] [Medline: 22123031]
27. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A virtual hope box: randomized controlled trial of a smartphone app for emotional regulation and coping with distress. Psychiatr Serv 2017 Apr 01;68(4):330-336. [doi: 10.1176/appi.ps.201600283] [Medline: 27842473]
28. 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]
29. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of SuperBetter, a smartphone-based/Internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
30. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
31. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
32. Buntrock C, Berking M, Smit F, Lehr D, Nobis S, Riper H, et al. Preventing depression in adults with subthreshold depression: health-economic evaluation alongside a pragmatic randomized controlled trial of a web-based intervention. J Med Internet Res 2017 Jan 04;19(1):e5 [FREE Full text] [doi: 10.2196/jmir.6587] [Medline: 28052841]
33. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
34. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
35. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
36. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
37. Furukawa TA, Horikoshi M, Kawakami N, Kadota M, Sasaki M, Sekiya Y, GENKI Project. Telephone cognitive-behavioral therapy for subthreshold depression and presenteeism in workplace: a randomized controlled trial. PLoS One 2012;7(4):e35330 [FREE Full text] [doi: 10.1371/journal.pone.0035330] [Medline: 22532849]
38. Takagaki K, Okamoto Y, Jinnin R, Mori A, Nishiyama Y, Yamamura T, et al. Behavioral activation for late adolescents with subthreshold depression: a randomized controlled trial. Eur Child Adolesc Psychiatry 2016 Nov;25(11):1171-1182. [doi: 10.1007/s00787-016-0842-5] [Medline: 27003390]
39. Simon GE, Ludman EJ, Tutty S, Operskalski B, Von Korff M. Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment: a randomized controlled trial. JAMA 2004 Aug 25;292(8):935-942. [doi: 10.1001/jama.292.8.935] [Medline: 15328325]
40. Watanabe N, Horikoshi M, Yamada M, Shimodera S, Akechi T, Miki K, Steering Committee of the Fun to Learn to ActThink through Technology Project. Adding smartphone-based cognitive-behavior therapy to pharmacotherapy for major depression

(FLATT project): study protocol for a randomized controlled trial. Trials 2015 Jul 07;16:293 [FREE Full text] [doi: 10.1186/s13063-015-0805-z] [Medline: 26149441]
41. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. Arlington, VA: American Psychiatric Association; 2013.
42. 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]
43. Beck A, Steer R, Brown G. BDI-II: Beck Depression Inventory. Second Edition. San Antonio, TX: The Psychological Corporation; 1996.
44. Thase ME, Rush AJ. When at first you don't succeed: sequential strategies for antidepressant nonresponders. J Clin Psychiatry 1997;58 Suppl 13:23-29 [FREE Full text] [Medline: 9402916]
45. 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]
46. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009 Feb 28;373(9665):746-758. [doi: 10.1016/S0140-6736(09)60046-5] [Medline: 19185342]
47. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [Medline: 10568646]
48. Furukawa TA. Assessment of mood: guides for clinicians. J Psychosom Res 2010 Jun;68(6):581-589. [doi: 10.1016/j.jpsychores.2009.05.003] [Medline: 20488276]
49. Rush AJ, Trivedi MH, Wisniewski SR, Stewart JW, Nierenberg AA, Thase ME, STAR*D Study Team. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. N Engl J Med 2006 Mar 23;354(12):1231-1242. [doi: 10.1056/NEJMoa052963] [Medline: 16554525]
50. Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. Control Clin Trials 2004 Apr;25(2):143-156. [doi: 10.1016/j.cct.2003.10.016] [Medline: 15020033]
51. Bang H. Random guess and wishful thinking are the best blinding scenarios. Contemp Clin Trials Commun 2016 Aug 15;3:117-121. [doi: 10.1016/j.conctc.2016.05.003] [Medline: 27822568]
52. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008 Jan 17;358(3):252-260. [doi: 10.1056/NEJMsa065779] [Medline: 18199864]
53. Hepner KA, Greenwood GL, Azocar F, Miranda J, Burnam MA. Usual care psychotherapy for depression in a large managed behavioral health organization. Adm Policy Ment Health 2010 May;37(3):270-278. [doi: 10.1007/s10488-009-0247-6] [Medline: 19859800]
54. Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. Acta Psychiatr Scand 2014 Sep;130(3):181-192. [doi: 10.1111/acps.12275] [Medline: 24697518]

Abbreviations

BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
FIBSER: Frequency, Intensity, and Burden of Side Effects Ratings
PHQ-9: Patient Health Questionnaire 9
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 29.07.17; peer-reviewed by K Hallgren, S Moritz; comments to author 24.08.17; revised version received 02.09.17; accepted 23.09.17; published 03.11.17.

Please cite as:
Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, Chino B, Funayama T, Yonemoto N, Zhou Q, Kawanishi N
Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial
J Med Internet Res 2017;19(11):e373
URL: http://www.jmir.org/2017/11/e373/
doi:10.2196/jmir.8602
PMID:29101095

Original Paper

Subregional Nowcasts of Seasonal Influenza Using Search Trends

Sasikiran Kandula[1], MS; Daniel Hsu[2], PhD; Jeffrey Shaman[1], PhD

[1]Department of Environmental Health Sciences, Columbia University, New York, NY, United States
[2]Department of Computer Science, Columbia University, New York, NY, United States

Corresponding Author:
Sasikiran Kandula, MS
Department of Environmental Health Sciences
Columbia University
ARB Building, 11th Floor
722 West 168th Street
New York, NY, 10032
United States
Phone: 1 2123053590
Fax: 1 2123054012
Email: sk3542@cumc.columbia.edu

Abstract

Background: Limiting the adverse effects of seasonal influenza outbreaks at state or city level requires close monitoring of localized outbreaks and reliable forecasts of their progression. Whereas forecasting models for influenza or influenza-like illness (ILI) are becoming increasingly available, their applicability to localized outbreaks is limited by the nonavailability of real-time observations of the current outbreak state at local scales. Surveillance data collected by various health departments are widely accepted as the reference standard for estimating the state of outbreaks, and in the absence of surveillance data, nowcast proxies built using Web-based activities such as search engine queries, tweets, and access of health-related webpages can be useful. Nowcast estimates of state and municipal ILI were previously published by Google Flu Trends (GFT); however, validations of these estimates were seldom reported.

Objective: The aim of this study was to develop and validate models to nowcast ILI at subregional geographic scales.

Methods: We built nowcast models based on autoregressive (autoregressive integrated moving average; ARIMA) and supervised regression methods (Random forests) at the US state level using regional weighted ILI and Web-based search activity derived from Google's Extended Trends application programming interface. We validated the performance of these methods using actual surveillance data for the 50 states across six seasons. We also built state-level nowcast models using state-level estimates of ILI and compared the accuracy of these estimates with the estimates of the regional models extrapolated to the state level and with the nowcast estimates published by GFT.

Results: Models built using regional ILI extrapolated to state level had a median correlation of 0.84 (interquartile range: 0.74-0.91) and a median root mean square error (RMSE) of 1.01 (IQR: 0.74-1.50), with noticeable variability across seasons and by state population size. Model forms that hypothesize the availability of timely state-level surveillance data show significantly lower errors of 0.83 (0.55-0.23). Compared with GFT, the latter model forms have lower errors but also lower correlation.

Conclusions: These results suggest that the proposed methods may be an alternative to the discontinued GFT and that further improvements in the quality of subregional nowcasts may require increased access to more finely resolved surveillance data.

(J Med Internet Res 2017;19(11):e370) doi:10.2196/jmir.7486

KEYWORDS

human influenza; classification and regression trees; nowcasts; infodemiology; infoveillance; surveillance

Introduction

Seasonal influenza infections are estimated to occur in 5% to 10% of the adult population worldwide annually, with higher attack rates in children and older adults [1,2]. In the United States, influenza accounts for about 1.2 deaths per 100,000 people, with considerable interseasonal variability [3]. Municipal and state health departments rely on surveillance data throughout the influenza season, typically October through May in the United States, to track the progress of the season and to

coordinate vaccination and treatment activities among hospitals, health care providers, and public health agencies. To support these efforts, the Centers for Disease Control and Prevention (CDC) disseminates weekly virologic and outpatient incidence data for influenza-like illness (ILI) at national and regional levels [4,5].

Several methods have been proposed to complement CDC's ILI, with estimates based on search queries [6-11], tweets [12,13], Wikipedia access logs [14,15], other public-generated content [16-18], and a combination of these proxies [19]. In addition to providing more timely estimates of outbreak progression, these data sources can be used to develop ILI estimates at the more localized subregional geographic resolutions at which public ILI data are limited or unavailable. As more effective and targeted interventions can be designed through a more local view of the outbreak, these subregional estimates, if accurate and reliable, are more actionable.

Google Flu Trends (GFT) [6] generated one of the more widely available estimates of ILI at regional and subregional levels using trends in Web-based search queries; however, production of GFT estimates was discontinued in August 2015 [20]. Instead, through Google's Extended Trends (GET) application programming interface (API), researchers are now permitted access to underlying Google search trends data and may build their own models to estimate ILI. The original GFT approach models CDC ILI as a linear function of search query frequency aggregated as a single variable. More recent work [7,21] demonstrated improved accuracy when individual query terms were retained as independent variables in the linear model, and further gain was reported with alternate models that allow for nonlinear and temporal relationships between the queries. A related study modeled ILI at week *w* on both autoregressive lags of n-weeks and search volume of 100 selected terms during week *w* [8,22].

Whereas these studies are encouraging, these models were developed and validated at US national level where the response variable, ILI, is available. Extrapolation of these national models to subregional resolutions where CDC ILI is not publicly available may yield nowcasts of limited accuracy. The GFT team is yet to publish the methodology used to generate nowcasts at subregional scales, and there have been few validation studies for GFT estimates at subregional levels [23,24].

In this paper, we propose methods to nowcast ILI at the subregional level using GET. These methods were applied retrospectively to generate nowcasts in 50 US states for six seasons, report the accuracy of different model forms, and compare these with GFT as published. It was observed that accurate nowcasts of subregional ILI may not be possible using models developed at the regional level; rather, subregional ILI nowcast models must be developed using subregional ILI data.

Methods

Overview

To build nowcast models at the US state level, random forest regression models were first built at the regional level (as defined by the US Department of Health and Human Services, HHS [25]). In these initial models, HHS regional weighted ILI, as reported by the CDC, was the response variable, and queries with search patterns correlated with ILI were explanatory variables. A 1-week ahead forecast of an autoregressive model fit on regional ILI was included as an additional explanatory variable. These regional-level models were then applied, or extrapolated, at the subregional scale. Specifically, the fit models were used with state-level explanatory variables to estimate ILI at the state level.

Independently, state-level nowcast models were built using CDC-provided state-level estimates of ILI as the response variable. These state-level ILI estimates are not publicly available and were provided for this study on request. The error of the state-level nowcast estimates made using these state models was then compared with the estimates of the regional models extrapolated to the state level.

Google Extended Trends (GET) Application Programming Interface

The GET API allows users to retrieve timeline data of the probability that a specified term is queried during a search session. Additional parameters allow specification of geographical (country, state, etc) and temporal (daily, weekly, etc) granularity and period of interest. Query probability is calculated on a random sample of 10% to 15% of all searches; terms whose search volume does not meet a minimum threshold are considered private, and their probabilities are reported as 0. Data updates are made daily and historical trends from January 2004 are available. Hence, nowcast models developed using GET can provide ILI estimates for at least one additional week over CDC ILI data, which are released with a 5- to 11-day lag.

In this study, as we were interested in state-level nowcasts, *state* was used as the geographical resolution and a *weekly* periodicity to be consistent with CDC ILI and GFT, both of which are weekly ILI estimates. We refer to logit transformed time series of term *t* as the query fraction of *t*, that is, $qf(t, s, w) = log(z/(1-z))$ where *z* is the probability that a query from state *s* during week *w* is for term *t*. GET does not provide separate query fractions at the HHS regional level. Therefore, the query fraction for a term from an HHS region was calculated as a population weighted mean of the query fractions for the term from states within the region. This choice of transformation was informed by previous work, which found that with logit transformation, the relation between raw query fractions and ILI becomes approximately linear and model performance improves [7].

Feature Identification

Queries highly correlated with CDC ILI were identified using Google Correlate [26,27] for use as explanatory variables. Google Correlate returns 100 queries whose search trends are historically most highly correlated (Pearson correlation coefficient) with a given target time series data. ILI at US national and 10 HHS regional levels from 2003-04 through 2014-15 influenza season was used as target time series. Significant overlap was observed in the queries identified using the different target time series. Query terms identified by Zhang

[28] and influenza-related entities extracted from Freebase [29], were appended to the list of correlates.

While examining the query fractions for terms related to ILI, it was found that some terms, which have considerable search activity at the national level, often have little-to-no activity at the state level and are reported as 0 (Multimedia Appendix 1; Figure S1), possibly because of the sampling and threshold criteria used in GET. Hence, the explanatory variables at the state level are sparse. To improve the quality of the data, a form of inheritance where a state inherits the query fraction of a term at the regional level when the state-level query fraction is zero was employed: *qf(t, s, w) = qf(t, r, w)*, where *s r*, and *r* designates HHS region. That is, in the absence of additional information, we assume users in all states of a region search for a term with the same probability. As this would not eliminate all zeros, the remaining zeros were replaced with a very small value, 1e-12, before applying the logit transformation. Sensitivity analysis showed that the results were not sensitive to the choice of the replacement (Multimedia Appendix 1; Figure S2).

Autoregressive Integrated Moving Average

Lampos et al [7] have found that simple autoregressive integrated moving average (ARIMA) models [30-32] using search trends data can generate reasonable nowcast estimates for ILI at the US national level. Similarly, Broniatowski et al [33,34] have demonstrated the utility of ARIMA models that use tweets and query data at a few subregional locations. ARIMA models are specified with three parameters, the order of the autoregressive component (a), the degree of differencing (d), and the order of the moving average component (q).

In Figure 1, ϕ , θ, and ρ are to be learned during model fitting. A method described by Hyndman and Khandakar [35,36] was used to search parameter space and to identify a set of parameters that provides good model fit, and the ARIMA models built at different times (*w*) and for different regions were allowed to use different parameters.

Figure 1. Autoregressive integrated moving average (ARIMA) formulation.

$$y_w = \sum_{i=1}^{a} \phi_i y_{w-i} + \sum_{i=1}^{d} \rho_i h_{w,i} + \sum_{i=1}^{q} \theta_i \varepsilon_{w-i} + \varepsilon_w$$

Random Forest

Random forest is a decision tree–based ensemble supervised learner that can be used for regression [37-39]. Specifically, given a dataset of *n* instances $D=(X, Y)=(x_{ip},y_i)$, where *Y* is a continuous response variable, and the feature set $X=(X_1, X_2,...,X_p)$ of *p* explanatory variables (ie, x_{ip} is the value of feature *j* for instance *i*), a supervised learning algorithm uses *D* to learn a function such that = *(X)* and minimizes some loss function with respect to *Y*. The function can then be used to estimate $_0$ for an instance $x_0=(x_{01}, x_{02}, \ldots,x_{op})$ whose response is unknown.

Decision tree–based methods split the feature space along an explanatory variable and learn separate fits, for each subspace. Ensemble methods build multiple decision trees, each tree on a dataset D^*, a random sampling with replacement of *n* instances from *D*. Random forests are ensemble decision trees that also exclude a random subset of explanatory variables while learning. Random forests are suitable for nonlinear problems with large feature sets and have been found to offer superior accuracy in multiple domains.

In this study, the randomForest [45] package in R [46] (R Project for Statistical Computing) was used to build the models.

Model Formulation

The model is described in detail in Multimedia Appendix 1. To summarize, let $y_{1:w}^{r}$ denote the logit transformed ILI observations for region *r* through week *w;* and $X_{1:v}^{r}$ a query fraction matrix of logit transformed query fractions at HHS region *r* for all terms in the feature set (columns) during weeks 1 through week *v* (rows). Note that *v>w*. We fit an ARIMA model on $y_{1:w}^{r}$ forecast ahead for weeks *w+1* to *v* and add the ARIMA result as an explanatory variable to $X_{1:v}^{r}$. With this modified matrix as the predictor and $(y_{1:w}^{r})^T$ as the response, we train a random forest model for region *r* at week *w*, $_w^r$. To nowcast ILI for a state *s* in region *r*, we append region *r*'s ARIMA results to the state's query fraction matrix $X_{1:v}^{s}$, and use this as a test set with $_w^r$.

Validation

State level ILI counts (per 100,000 patient visits) from 2000-01 to 2010-11 season were provided by CDC following a data request. These counts were considered as the true values to validate the estimates from the model described above. As GET data were only available from January 2004, the last six of the seven overlapping flu seasons (*Morbidity and Mortality Weekly Report* [40], MMWR, week 40 through MMWR week 39 of the next calendar year), that is, 2005-06 to 2010-11 were used for validation. To generate nowcasts for any given week, only data that would have been available if nowcasts were being generated in real time were used, thus allowing for an out-of-sample validation of the estimates.

For each state during each of the six seasons, the Pearson correlation coefficient (COR), root mean square error (RMSE), and mean absolute proportion error (MAPE) were calculated. In Figure 2, y_w^s is the true ILI for state *s* at week *w*, $_w^s$ the corresponding nowcast, *w se* the weeks in a given flu season, and *g()* is the inverse logit transformation. Although nowcast estimates up to 2 weeks ahead are sometimes possible using ARIMA and GET, only 1-week ahead estimates were used in this error analysis.

Figure 2. Formulation for two error measures: root mean square error (RMSE) and mean absolute proportion error (MAPE).

$$RMSE_{se}^{s} = \sqrt{\frac{1}{|se|}\sum_{w \in se}\left(g(\hat{y}_w^s) - g(y_w^s)\right)^2}$$

$$MAPE_{se}^{s} = \frac{1}{|se|}\sum_{w \in se}\frac{|g(\hat{y}_w^s) - g(y_w^s)|}{g(y_w^s)}$$

Alternate Model Forms

To generate nowcasts for a state, the model trained with its corresponding regional data was extrapolated to the state level. For this extrapolation, the model formulation, described above and trained using regional ILI as the response variable, was applied using an ARIMA fit of regional ILI and state GET query fractions as explanatory variables. We refer to this form as RRS. Two other alternate forms were explored: RR0, where the state's ILI estimate is simply its region's ARIMA estimate, and RRR, where the state's GET query fractions were replaced with the query fractions for its parent region.

The accuracy of RRS relative to RR0 was indicative of value added by GET and random forests, and of RRS relative to RRR as value added through the use of more localized GET data. As GFT was published during the six seasons being used for validation, the performance of these three model forms were also compared against GFT.

Alternate Model Forms: State ILI as Response

The three model forms described above were built with regional ILI as the response variable. As regional ILI is released weekly by the CDC, these models are suitable for real-time operational nowcasts. Although estimates of subregional ILI are not publicly available, state and municipal health agencies do have these estimates for internal use, and it is worthwhile to develop and test model forms that would be possible with subregional ILI.

Four additional model forms were defined: SS0, a simple ARIMA model fit on state ILI; SRR and SRS, which are similar to RRR and RRS, respectively, except for the response variable used for training; and SSS, which does not directly use any regional information. Please see Multimedia Appendix 1 for more formal specification of these four types.

To compare the different model forms and to check that the differences were statistically significant, we used a Friedman rank-sum test [41,42] followed by a Nemenyi test [43,44]. The Friedman test is a nonparametric test that does not assume normality. It ranks the different model forms on each test attempt, a state-season combination and uses the rank to test whether model forms are different. The Nemenyi test, a post hoc test for Friedman, checks for statistically significant differences between each pair of model forms.

Results

Of the explanatory variables used in the RRS models, the ARIMA component (*ar*) ranks highest followed by a good number of entities from Freebase (see Figure 3). Across all seasons and states, the RRS models were found to have a reasonably high median correlation of 0.84 (interquartile range [IQR]: 0.74-0.91; Table 1). When stratified by population size, states with larger population sizes had significantly higher median correlations than those with small population sizes. Significant variability across seasons was also observed. States with large populations sizes were also found to have lower median errors (RMSE and MAPE), but there does not seem to be much difference between low- and medium-sized states.

Although the correlation of the RRS models was encouraging, GFT estimates have better median measures overall and across almost all disaggregated groups. Google has not published their method to estimate ILI at subregional levels, and it is not clear whether GFT estimates benefited from a fuller access to trends data or whether the performance gain was solely methodological.

Figure 3. Top 20 features by importance as determined by random forest models built at regional level. The dot and whiskers in red show the median and interquartile range (IQR), respectively, whereas the blue point is the mean. The label shows the percentage of models in which the feature was used (n=3130). ar refers to the autoregressive integrated moving average (ARIMA) component. Features prefixed by ENT are entities identified using Freebase.

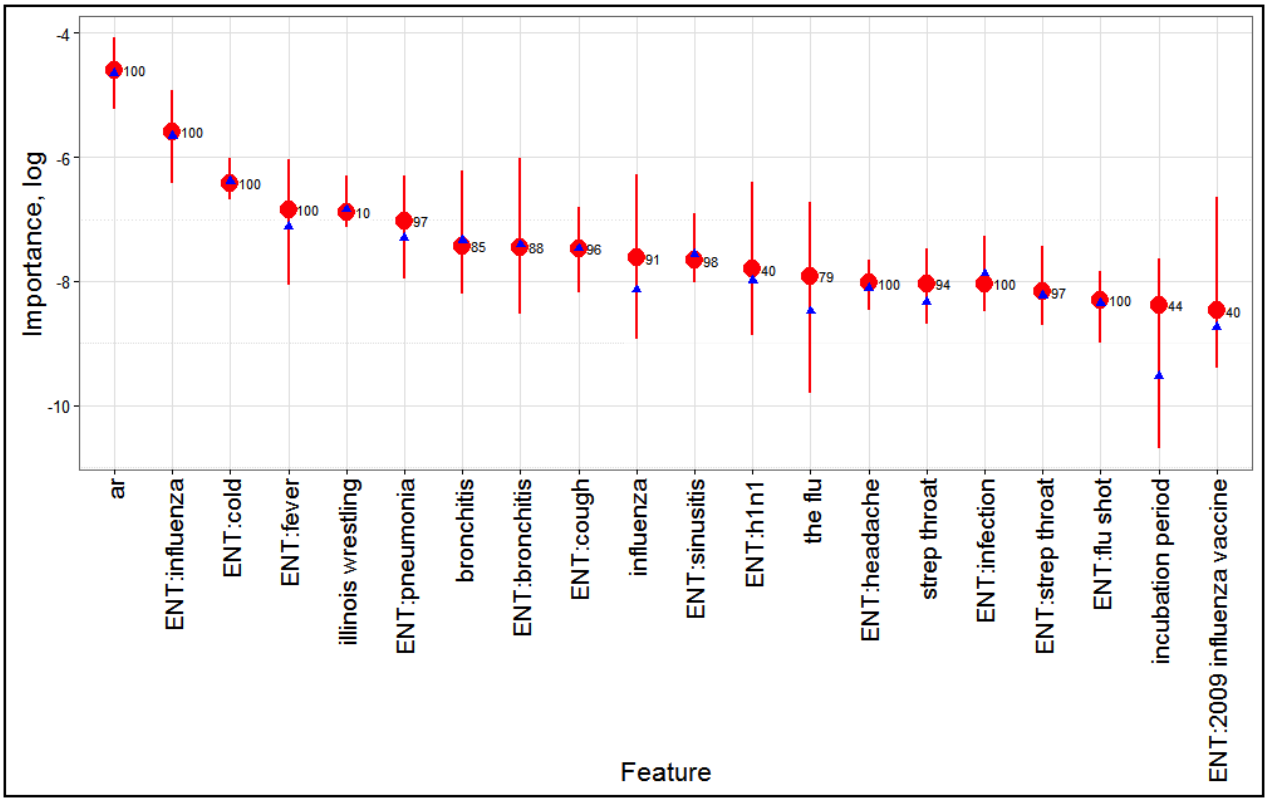


Table 1. Median (interquartile range), Pearson correlation coefficient (COR), root mean square error (RMSE), and mean absolute proportion error (MAPE) for RRS, RR0, RRR models, and Google Flu Trends (GFT). Results are stratified by state population size and season.

Measure	RRS, median (interquartile range)	RR0, median (interquartile range)	RRR, median (interquartile range)	GFT[a], median (interquartile range)
COR [b]				
Overall	0.85 (0.74-0.91)	0.83 (0.7-0.9)	0.86 (0.75-0.91)	0.89 (0.8-0.94)
Population size (millions)				
0-2 (n=14)	0.79 (0.64-0.87)	0.76 (0.62-0.86)	0.81 (0.67-0.88)	0.83 (0.72-0.91)
2-5 (n=14)	0.84 (0.72-0.89)	0.82 (0.7-0.89)	0.84 (0.75-0.90)	0.9 (0.81-0.94)
5-7.5 (n=10)	0.84 (0.74-0.91)	0.82 (0.7-0.9)	0.86 (0.73-0.92)	0.89 (0.8-0.95)
≥7.5 (n=12)	0.91 (0.85-0.93)	0.9 (0.84-0.93)	0.91 (0.86-0.94)	0.93 (0.86-0.96)
Season				
05-06	0.8 (0.62-0.85)	0.8 (0.62-0.85)	0.81 (0.64-0.87)	0.83 (0.71-0.88)
06-07	0.82 (0.65-0.88)	0.8 (0.6-0.88)	0.82 (0.71-0.89)	0.83 (0.76-0.9)
07-08	0.88 (0.81-0.92)	0.87 (0.79-0.92)	0.89 (0.82-0.93)	0.93 (0.87-0.96)
08-09	0.75 (0.69-0.83)	0.71 (0.58-0.82)	0.78 (0.67-0.83)	0.81 (0.71-0.89)
09-10	0.9 (0.85-0.93)	0.89 (0.8-0.93)	0.9 (0.85-0.93)	0.97 (0.94-0.98)
10-11	0.89 (0.82-0.92)	0.88 (0.75-0.91)	0.89 (0.85-0.92)	0.89 (0.86-0.93)
RMSE [c]				
Overall	0.99 (0.7-1.51)	1.06 (0.73-1.56)	0.97 (0.72-1.54)	0.93 (0.66-1.33)
Population size (millions)				
0-2 (n=14)	1.06 (0.69-1.58)	1.19 (0.73-1.62)	1.05 (0.72-1.6)	0.88 (0.63-1.29)
2-5 (n=14)	1.21 (0.84-1.87)	1.33 (0.92-1.81)	1.22 (0.83-1.84)	1.02 (0.78-1.52)
5-7.5 (n=10)	0.93 (0.65-1.21)	0.98 (0.72-1.33)	0.93 (0.61-1.14)	0.88 (0.67-1.48)
≥7.5 (n=12)	0.87 (0.66-1.01)	0.85 (0.70-1.08)	0.88 (0.69-1.01)	0.87 (0.63-1.16)
Season				
05-06	0.93 (0.64-1.5)	0.92 (0.70-1.64)	0.93 (0.64-1.52)	0.88 (0.60-1.45)
06-07	0.84 (0.56-1.16)	0.89 (0.57-1.16)	0.85 (0.5-1.1)	0.82 (0.52-1.13)
07-08	1.08 (0.81-1.7)	1.06 (0.83-1.59)	0.99 (0.82-1.67)	1.09 (0.70-1.55)
08-09	1.02 (0.77-1.47)	1.10 (0.79-1.48)	1.03 (0.79-1.55)	1.02 (0.79-1.41)
09-10	1.31 (0.98-1.77)	1.40 (1.08-1.72)	1.28 (0.98-1.72)	1.05 (0.80-1.32)
10-11	0.77 (0.59-1.16)	0.83 (0.61-1.26)	0.83 (0.59-1.15)	0.73 (0.64-1.20)
MAPE [d] **(/1000)**				
Overall	0.8 (0.43-1.75)	0.67 (0.42-1.54)	0.77 (0.43-1.62)	0.71 (0.44-1.51)
Population size (millions)				
0-2 (n=14)	0.9 (0.54-1.7)	0.77 (0.51-1.41)	0.84 (0.55-1.55)	0.76 (0.51-1.56)
2-5 (n=14)	0.95 (0.48-1.79)	0.82 (0.44-1.65)	0.87 (0.45-1.71)	0.77 (0.41-1.48)
5-7.5 (n=10)	0.65 (0.36-1.62)	0.59 (0.37-1.69)	0.63 (0.35-1.57)	0.68 (0.4-1.41)
≥7.5 (n=12)	0.65 (0.34-1.64)	0.54 (0.3-1.34)	0.65 (0.33-1.5)	0.7 (0.43-1.54)
Season				
05-06	1.2 (0.46-3.06)	0.78 (0.47-2.77)	0.99 (0.49-2.72)	1.07 (0.56-2.67)
06-07	0.97 (0.53-1.84)	0.92 (0.49-1.81)	0.91 (0.51-1.67)	0.88 (0.46-1.48)
07-08	0.85 (0.5-1.67)	0.83 (0.49-1.64)	0.81 (0.51-1.51)	0.76 (0.5-1.57)
08-09	0.82 (0.47-1.59)	0.67 (0.43-1.36)	0.84 (0.43-1.52)	0.71 (0.44-1.48)

Measure	RRS, median (interquartile range)	RR0, median (interquartile range)	RRR, median (interquartile range)	GFT[a], median (interquartile range)
09-10	0.73 (0.36-1.96)	0.64 (0.4-1.83)	0.74 (0.36-1.96)	0.63 (0.43-1.17)
10-11	0.49 (0.3-1.04)	0.48 (0.28-0.96)	0.48 (0.31-1.04)	0.61 (0.32-0.93)

[a]GFT: Google Flu Trends.

[b]COR: Pearson correlation coefficient.

[c]RMSE: root mean square error.

[d]MAPE: mean absolute percentage error.

Table 2. Mean rank and statistical significance from post hoc Nemenyi test. For each season-state combination, the model forms are ranked from best (rank=1) to worst (rank=4).

Model	COR[a]				RMSE[b]				MAPE[c]			
	Mean rank	GFT[d]	RRO	RRR	Mean rank	GFT	RRO	RRR	Mean rank	GFT	RRO	RRR
GFT	1.91				2.33				2.45			
RR0	3.07	<.001			2.75	<.001			2.24	.17		
RRR	2.38	<.001	<.001		2.41	.89	.01		2.43	.99	.25	
RRS	2.63	<.001	<.001	.1	2.51	.35	.09	.79	2.87	<.001	<.001	<.001

[a]COR: Pearson correlation coefficient.

[b]RMSE: root mean square error.

[c]MAPE: mean absolute percentage error.

[d]GFT: Google Flu Trends.

Table 2 shows the mean rank for the model forms along with the results of Friedman-Nemenyi tests for significance. Of the four estimates, the best performing (highest correlation or lowest error) is assigned a rank of 1, the worst a rank of 4, and an average across the different season-state combinations (n=300) is calculated. The results indicate the following: (1) For correlation, GFT has the best mean rank, followed by RRR, RRS, and RR0. However, the difference between RRR and RRS is not statistically significant; (2) the relative ordering of the mean ranks remains the same with RMSE, but the difference between RR0, RRR, and RRS is not statistically significant; and, (3) RR0 has the best rank with MAPE followed by GFT. The mean ranks of RRR and RRS are significantly higher.

Overall, the performance of the RRR models was comparable to the RRS models, which indicates that state-localized GET data, as used in the models described here, do not improve nowcast accuracy. Because RR0 lowers (degrades) correlation, does not alter RMSE and considerably lowers (improves) MAPE, the effect of ignoring GET data altogether remains uncertain.

Extending the comparison to model forms that are built using state ILI as the response variable (Table 3; Figures 4 and 5), a noticeable reduction was observed in errors. The median RMSE and MAPE (Figure 4) of the SRS, SRR, and SSS models are lower than GFT overall, in states with larger population, and in a majority of the seasons. There is also a clear improvement over their RR* counterparts (Figure 5). However, the median correlation of all four models is noticeably lower, especially for the SS0 models.

Table 3. Median (interquartile range), Pearson correlation coefficient (COR), root mean square error (RMSE), and mean absolute percentage error (MAPE) for Google Flu Trends (GFT), SS0, SRR, SRS, and SSS models. Results are stratified by state population and season.

Measure	GFT[a], median (interquartile range)	SS0, median (interquartile range)	SRR, median (interquartile range)	SRS, median (interquartile range)	SSS, median (interquartile range)
COR [b]					
Overall	0.89 (0.8-0.94)	0.56 (0.4-0.75)	0.8 (0.7-0.88)	0.8 (0.7-0.88)	0.74 (0.61-0.83)
Population size (millions)					
0-2 (n=14)	0.83 (0.72-0.91)	0.46 (0.31-0.66)	0.74 (0.57-0.82)	0.71 (0.56-0. 8)	0.62 (0.55-0.74)
2-5 (n=14)	0.9 (0.81-0.94)	0.58 (0.42-0.76)	0.78 (0.72-0.87)	0.8 (0.72-0.85)	0.73 (0.66-0.81)
5-7.5 (n=10)	0.89 (0.8-0.95)	0.51 (0.36-0.64)	0.83 (0.7-0.88)	0.81 (0.73-0.88)	0.75 (0.63-0.82)
≥7.5 (n=12)	0.93 (0.86-0.96)	0.73 (0.48-0.85)	0.88 (0.79-0.92)	0.89 (0.8-0.92)	0.86 (0.72-0.91)
Season					
05-06	0.83 (0.71-0.88)	0.72 (0.56-0.85)	0.78 (0.68-0.86)	0.76 (0.62-0.86)	0.74 (0.66-0.86)
06-07	0.83 (0.76-0.9)	0.75 (0.61-0.84)	0.8 (0.7-0.88)	0.8 (0.64-0.87)	0.8 (0.72-0.89)
07-08	0.93 (0.87-0.96)	0.61 (0.47-0.77)	0.87 (0.78-0.92)	0.86 (0.78-0.9)	0.81 (0.73-0.86)
08-09	0.81 (0.71-0.89)	0.37 (0.28-0.44)	0.7 (0.59-0.8)	0.74 (0.58-0.79)	0.57 (0.45-0.68)
09-10	0.97 (0.94-0.98)	0.51 (0.39-0.73)	0.82 (0.75-0.89)	0.82 (0.74-0.89)	0.74 (0.63-0.85)
10-11	0.89 (0.86-0.93)	0.47 (0.33-0.6)	0.82 (0.75-0.88)	0.81 (0.75-0.88)	0.71 (0.63-0.78)
RMSE [c] **(1e-3)**					
Overall	0.93 (0.66-1.33)	1.07 (0.68-1.84)	0.84 (0.54-1.25)	0.86 (0.55-1.27)	0.9 (0.55-1.35)
Population size (millions)					
0-2 (n=14)	0.88 (0.63-1.29)	1.17 (0.61-1.92)	0.96 (0.55-1.47)	0.96 (0.62-1.49)	0.92 (0.58-1.44)
2-5 (n=14)	1.02 (0.78-1.52)	1.37 (0.83-2.13)	1.04 (0.7-1.54)	1.11 (0.62-1.57)	1.11 (0.66-1.68)
5-7.5 (n=10)	0.88 (0.67-1.48)	0.99 (0.66-1.79)	0.74 (0.49-1.07)	0.71 (0.51-1.14)	0.79 (0.55-1.24)
≥7.5 (n=12)	0.87 (0.63-1.16)	0.91 (0.64-1.49)	0.69 (0.43-1.05)	0.67 (0.41-0.99)	0.74 (0.46-1.01)
Season					
05-06	0.88 (0.60-1.45)	0.81 (0.49-1.47)	0.71 (0.5-1.11)	0.68 (0.49-1.13)	0.64 (0.46-1.06)
06-07	0.82 (0.52-1.13)	0.70 (0.48-1.02)	0.59 (0.43-0.88)	0.58 (0.42-0.94)	0.56 (0.41-0.83)
07-08	1.09 (0.70-1.55)	1.36 (0.78-1.85)	0.91 (0.54-1.27)	0.95 (0.58-1.37)	0.97 (0.6-1.42)
08-09	1.02 (0.79-1.41)	1.21 (0.92-1.98)	0.95 (0.69-1.31)	0.93 (0.67-1.26)	1.05 (0.78-1.4)
09-10	1.05 (0.80-1.32)	1.91 (1.28-2.44)	1.34 (0.9-1.9)	1.37 (0.92-1.92)	1.53 (1.01-1.9)
10-11	0.73 (0.64-1.20)	1.00 (0.73-1.62)	0.73 (0.5-1.04)	0.7 (0.51-1.1)	0.86 (0.58-1.16)
MAPE [d]					
Overall	0.71 (0.44-1.51)	0.58 (0.38-0.8)	0.54 (0.33-0.9)	0.61 (0.34-1)	0.61 (0.35-1.02)
Population size (millions)					
0-2 (n=14)	0.76 (0.51-1.56)	0.68 (0.48-0.86)	0.76 (0.5-1.36)	0.84 (0.56-1.44)	0.82 (0.58-1.28)
2-5 (n=14)	0.77 (0.41-1.48)	0.63 (0.36-0.85)	0.58 (0.36-0.9)	0.64 (0.39-1)	0.68 (0.37-1.02)
5-7.5 (n=10)	0.68 (0.4-1.41)	0.58 (0.39-0.74)	0.41 (0.31-0.75)	0.46 (0.32-0.86)	0.55 (0.34-0.92)
≥7.5 (n=12)	0.7 (0.43-1.54)	0.4 (0.31-0.59)	0.38 (0.2-0.59)	0.37 (0.2-0.69)	0.41 (0.24-0.61)
Season					
05-06	1.07 (0.56-2.67)	0.59 (0.39-0.8)	0.68 (0.4-0.93)	0.77 (0.41-1.12)	0.74 (0.38-1.08)
06-07	0.88 (0.46-1.48)	0.54 (0.36-0.71)	0.51 (0.32-0.84)	0.62 (0.35-0.94)	0.58 (0.3-0.89)
07-08	0.76 (0.5-1.57)	0.69 (0.4-0.83)	0.54 (0.38-0.78)	0.62 (0.41-0.94)	0.62 (0.38-0.81)
08-09	0.71 (0.44-1.48)	0.57 (0.42-0.77)	0.62 (0.37-1.01)	0.66 (0.36-0.93)	0.68 (0.39-1.14)

Measure	GFT[a], median (interquartile range)	SS0, median (interquartile range)	SRR, median (interquartile range)	SRS, median (interquartile range)	SSS, median (interquartile range)
09-10	0.63 (0.43-1.17)	0.59 (0.36-0.85)	0.52 (0.31-1.25)	0.59 (0.31-1.38)	0.67 (0.37-1.14)
10-11	0.61 (0.32-0.93)	0.5 (0.35-0.85)	0.38 (0.26-0.67)	0.38 (0.26-0.75)	0.43 (0.31-0.83)

[a]GFT: Google Flu Trends.

[b]COR: Pearson correlation coefficient.

[c]RMSE: root mean square error.

[d]MAPE: mean absolute percentage error.

Figure 4. Measures observed with the different model forms A: Pearson correlation coefficient (COR); B: Root mean square error (RMSE); and C: Mean absolute percentage error (MAPE). Left: The box and whiskers show the median, interquartile range (IQR), and extrema (1.5×IQR) for each model form. The colored regions are violin plots showing probability density. Right: Heat map of the distribution of relative ranks of the models; more frequent ranks are darker.

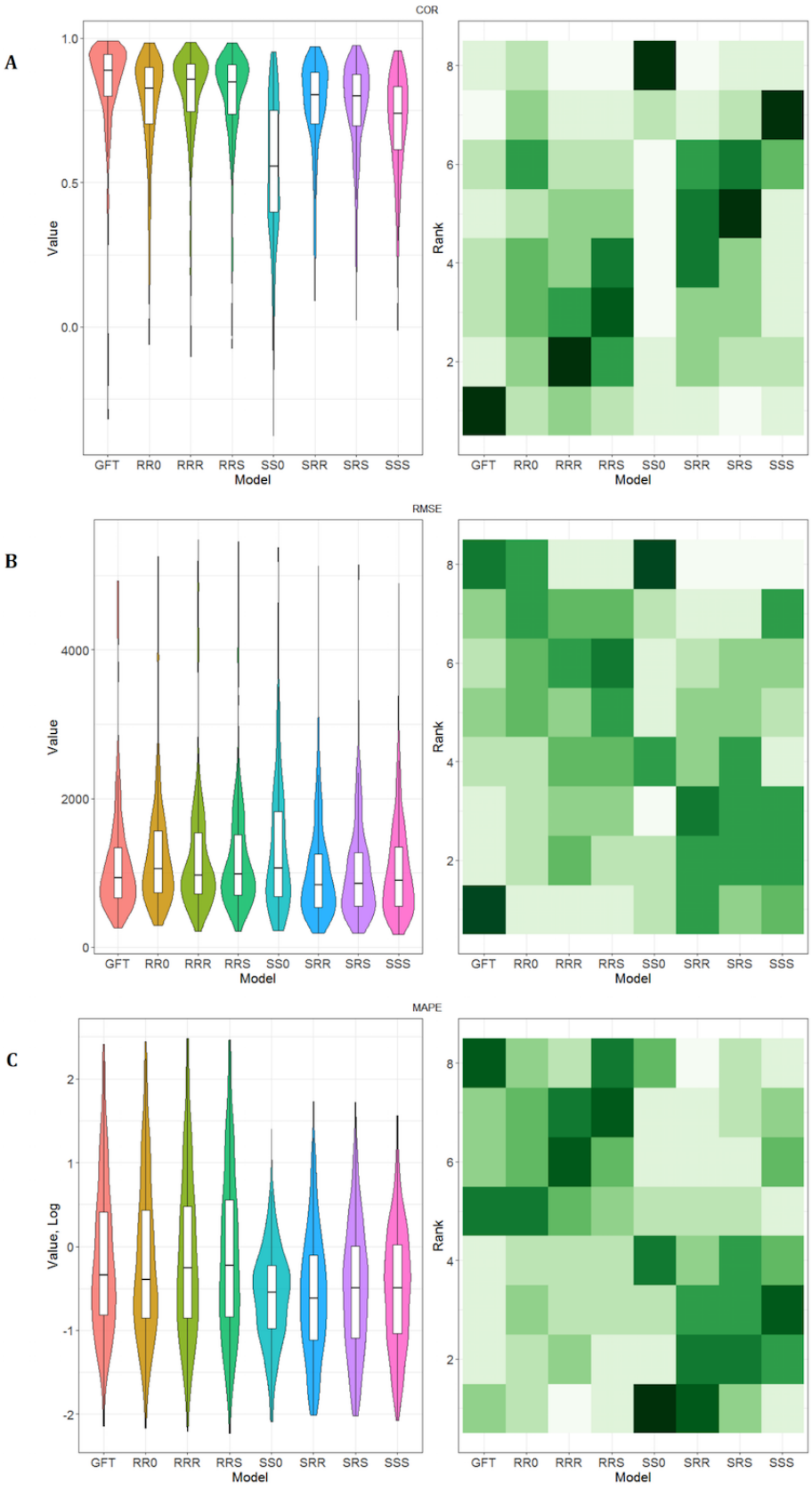


Figure 5. Pairwise plots for the model forms on the three measures forms A: Pearson correlation coefficient (COR); B: Root mean square error (RMSE); and C: Mean absolute percentage error (MAPE). The subpanels along the diagonal show density of the measure for the model form. Subpanels in the lower triangle are scatter plots (n=300) denoting a state-season. Points on or close to the black line (y=x) are state-seasons where the pair of model forms have similar measures (correlation or error). Subpanels in the upper triangle are heat maps of the counts of points in each two-dimensional (2D) grid of the plot area (low counts in yellow, high in red). For example, to compare the correlations of RRS and SS0, see the scatter plot in (5,4) or heat map in (4,5) of A.

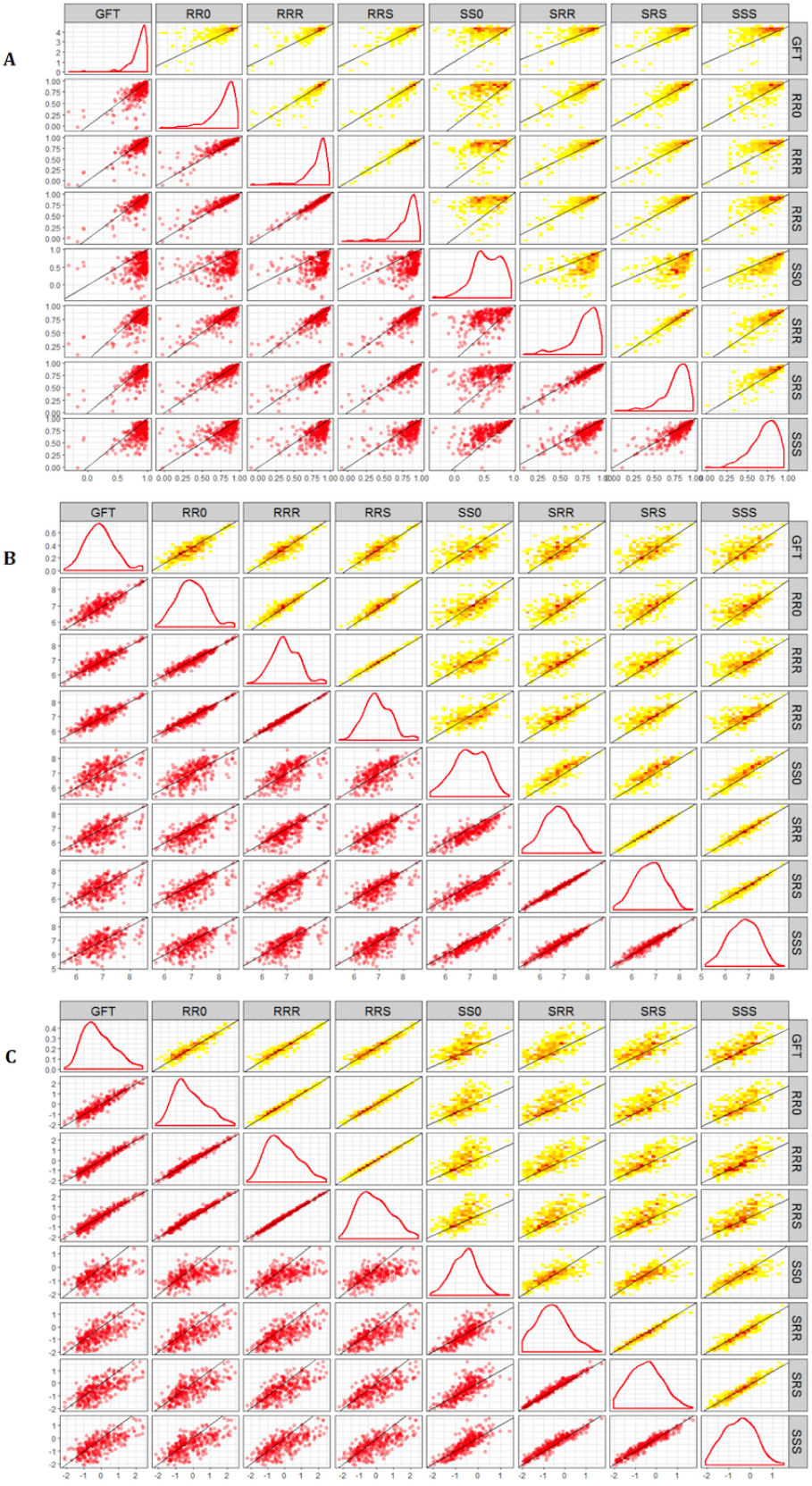


Table 4. Mean rank and statistical significance from post hoc Nemenyi test. For each season-state combination, the model forms are ranked from best (rank=1) to worst (rank=8).

Measure	Model	Mean rank	GFT[a]	RRO	RRR	RRS	SS0	SRR	SRS
Pearson correlation coefficient (COR)	GFT	2.67							
	RR0	4.55	<.001						
	RRR	3.34	.002	<.001					
	RRS	3.68	<.001	<.001	.68				
	SS0	6.87	<.001	<.001	<.001	<.001			
	SRR	4.37	<.001	.98	<.001	.01	<.001		
	SRS	4.75	<.001	.97	<.001	<.001	<.001	.55	
	SSS	5.73	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Root mean square error (RMSE)	GFT	4.46							
	RR0	5.27	.002						
	RRR	4.68	.96	.06					
	RRS	4.82	.62	.35	.99				
	SS0	5.77	<.001	.19	<.001	<.001			
	SRR	3.34	<.001	<.001	<.001	<.001	<.001		
	SRS	3.71	.005	<.001	<.001	<.001	<.001	.61	
	SSS	3.96	.2	<.001	<.001	<.001	<.001	.04	.92
Mean absolute proportion error (MAPE)	GFT	5.26							
	RR0	4.91	.65						
	RRR	5.18	.99	.89					
	RRS	5.7	.37	.002	.15				
	SS0	3.75	<.001	<.001	<.001	<.001			
	SRR	3.17	<.001	<.001	<.001	<.001	.07		
	SRS	3.93	<.001	<.001	<.001	<.001	.99	<.001	
	SSS	4.09	<.001	.001	<.001	<.001	.69	<.001	.99

[a]GFT: Google Flu Trends.

Results from Friedman-Nemenyi tests (see Table 4) show that SRS has the lowest mean rank for RMSE, and the difference is statistically significant from all other models, with the exception of SRR. SS0 has the lowest mean rank for MAPE but is not statistically different from either SRS or SRR. It is also interesting to note that models that continue to use ARIMA fit on regional ILI (SRR and SRS) match or outperform those that use ARIMA fit on state ILI (SS0 and SSS).

Discussion

Principal Findings

We described a method to nowcast ILI at subregional levels using GET and validated the developed models against real surveillance data across six influenza seasons and 50 states in the United States. The method was found to give improved estimates over an autoregressive model but underperformed relative to GFT. Variants of the method that used surveillance data at subregional levels, in a majority of the cases, bettered GFT.

Our results support earlier findings by other groups of the suitability of ARIMA models, both by themselves and in conjunction with other methods, in nowcasting ILI. This has particular relevance for very small settings, say a hospital or a rural county health department, where internal estimates of ILI are available and short-horizon forecasts are of interest for resource planning.

It was also found that data accessible through GET API are sparse at finer geographical granularity, and methods that rely solely on search trend data may not be viable for localized nowcasts. The inheritance method described here addresses the issue to some extent, as tests for the impact of inheritance on the models' performance found that inheritance improves correlation overall, particularly in states with low population; however, it has no significant impact on RMSE and increases MAPE (Multimedia Appendix 1; Figure S3). Additional analysis is necessary to identify scenarios, for example, when a state's signal is below a fraction of the parent region or below a threshold determined by historical likelihood, in which inheritance is useful. Incorporation of alternate data

streams—such as electronic health records and social media—as additional features to the random forest models may obviate the need for inheritance and potentially improve nowcasts.

The reduced errors of the S* models, which use state-level ILI as the training response variable, make a case for the public release of this information every week. CDC estimates ILI at HHS regions by aggregating data submitted through the US Outpatient ILI Surveillance Network (ILINet) by about 2000 outpatient health care providers in the United States every week. Aggregation of data at subregional levels is possible in theory, but there are concerns about patient and provider privacy. However, given our findings that reliance on regional ILI with or without subregional GET produces inferior subregional nowcasts and that these are only marginally better than use of regional ILI as a proxy for subregional ILI, perhaps it is necessary to revisit specific concerns about privacy and to explore anonymization methods whose use might permit release of ILINet data at the subregional level.

As all states in an HHS region will have the same RRR nowcast estimate, the performance of RRR and GFT in nowcasting *regional* ILI can be compared. No significant difference was found between RRR nowcasts and GFT at the regional level for any of the three accuracy measures used (see Multimedia Appendix 1; Table S4). The superior performance of GFT over R* models at the state level, however, requires additional analysis. Although we have little information on the GFT model form, we believe that Google had no access to subregional CDC ILI data to train subregional models. As a consequence, GFT municipal- and state-level ILI estimates were likely extrapolations of regional models, akin to the R* models described here. This might also explain why our S* models outperform GFT in terms of RMSE and MAPE—by building models at the state level, biases in state level ILI data relative to the parent region were eliminated, thereby reducing error (this implicit bias correction is indeed observed; see Multimedia Appendix 1; Figure S4). If GFT had the same access to search trends as is now publicly available through GET, the superior GFT subregional nowcasts relative to R* models suggest that both the feature set and the learning method presented here need to be improved further. If, on the other hand, GFT had full (100%) access to GET, then its superior performance relative to R* models may stem more from that discrepancy in access.

One limitation of the validation method reported above is that it does not account for back-revisions to ILI data. CDC's ILI estimates are updated for multiple weeks following the week of initial release, as additional providers submit delayed data. We did not have access to information on how state-level ILI was updated over time but only to the final stable ILI. If this detailed versioned dataset were available, a more robust validation comparing nowcasts generated using transient estimates of ILI with the final stable ILI would have been possible.

Conclusions

Overall, the findings suggest that nowcast extrapolation to more local scales are likely to remain challenging, as long as data at these scales remain restricted. As public health interventions and hospital planning can benefit from timely and localized estimates of ILI, relaxation of these restrictions may be warranted.

Acknowledgments

This work was supported by grants from the US National Institutes of Health (NIH; GM110748 to JS and SK; GM100467 to JS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors would like to thank Christian Stefansen and Google Health Trends team for useful discussions and help with the API and data and Mehmet Turkcan for collaborating in the development of earlier versions of some of the model forms.

Conflicts of Interest

JS declares partial ownership of SK Analytics. SK was a contractor for SK Analytics.

Multimedia Appendix 1

Supporting information.

[PDF File (Adobe PDF File), 1MB - jmir_v19i11e370_app1.pdf]

References

1. WHO. Influenza (seasonal) fact sheet URL: http://www.who.int/mediacentre/factsheets/fs211/en/ [accessed 2017-09-04] [WebCite Cache ID 6tEctpQxS]
2. WHO. Influenza vaccines URL: http://www.who.int/biologicals/vaccines/influenza/en/ [accessed 2017-09-04] [WebCite Cache ID 6tEcxKmG9]
3. Xu J, Murphy SL, Kochanek KD, Bastian BA. Deaths: final data for 2013. Natl Vital Stat Rep 2016;64(2):1-119 [FREE Full text] [Medline: 26905861]
4. CDC. Overview of influenza surveillance in the United States URL: http://www.cdc.gov/flu/weekly/overview.htm [accessed 2017-09-04] [WebCite Cache ID 6tEd2Ix1L]

5. CDC. FluView interactive URL: https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html [accessed 2017-09-04] [WebCite Cache ID 6tEd6aocm]
6. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
7. Lampos V, Miller AC, Crossan S, Stefansen C. Advances in nowcasting influenza-like illness rates using search query logs. Sci Rep 2015 Aug 03;5:12760 [FREE Full text] [doi: 10.1038/srep12760] [Medline: 26234783]
8. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci U S A 2015 Nov 24;112(47):14473-14478 [FREE Full text] [doi: 10.1073/pnas.1515373112] [Medline: 26553980]
9. Eysenbach G, Köhler C. Health-related searches on the Internet. J Am Med Assoc 2004 Jun 23;291(24):2946. [doi: 10.1001/jama.291.24.2946] [Medline: 15213205]
10. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
11. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 01;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
12. Dredze M, Paul MJ, Bergsma S, Tran H. Carmen: A twitter geolocation system with applications to public health. 2013 Presented at: AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI); July 14-18, 2013; Bellevue, Washington, USA.
13. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014 Oct 28;6:1-2 [FREE Full text] [doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117] [Medline: 25642377]
14. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol 2014 Apr;10(4):e1003581 [FREE Full text] [doi: 10.1371/journal.pcbi.1003581] [Medline: 24743682]
15. Hickmann KS, Fairchild G, Priedhorsky R, Generous N, Hyman JM, Deshpande A, et al. Forecasting the 2013-2014 influenza season using Wikipedia. PLoS Comput Biol 2015 May;11(5):e1004239 [FREE Full text] [doi: 10.1371/journal.pcbi.1004239] [Medline: 25974758]
16. Ray J, Brownstein J. Nowcasting influenza activity using Healthmap data. 2015 Presented at: DTRA Chemical Biological Defense Conference; May 12-14, 2015; St. Louis, MO URL: https://www.osti.gov/scitech/servlets/purl/1251371
17. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu near you: Crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015 Oct;105(10):2124-2130. [doi: 10.2105/AJPH.2015.302696] [Medline: 26270299]
18. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 2015 Oct;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
19. Farrow D. 2016. Modeling the past, present, and future of influenza URL: https://delphi.midas.cs.cmu.edu/~dfarrow/thesis.pdf [accessed 2017-10-08] [WebCite Cache ID 6u3PQq9jv]
20. Google Research Blog. The next chapter for flu trends URL: https://research.googleblog.com/2015/08/the-next-chapter-for-flu-trends.html [accessed 2017-09-04] [WebCite Cache ID 6tEdDvoZJ]
21. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to) Google Flu Trends? Am J Prev Med 2014 Sep;47(3):341-347. [doi: 10.1016/j.amepre.2014.05.020] [Medline: 24997572]
22. Tibshirani R. 1996. Regression shrinkage and selection via the lasso URL: https://statweb.stanford.edu/~tibs/lasso/lasso.pdf [accessed 2017-10-08] [WebCite Cache ID 6u3Pw8ZbB]
23. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol 2013;9(10):e1003256 [FREE Full text] [doi: 10.1371/journal.pcbi.1003256] [Medline: 24146603]
24. Pollett S, Boscardin WJ, Azziz-Baumgartner E, Tinoco YO, Soto G, Romero C, et al. Evaluating Google Flu Trends in Latin America: important lessons for the next phase of digital disease detection. Clin Infect Dis 2017 Jan 01;64(1):34-41. [doi: 10.1093/cid/ciw657] [Medline: 27678084]
25. U.S. Department of Health & Human Services. HHS regional offices. URL: https://www.hhs.gov/about/agencies/regional-offices/index.html [accessed 2017-09-04] [WebCite Cache ID 6tEcpnQ0c]
26. Google. Google correlate URL: https://www.google.com/trends/correlate [accessed 2017-09-04] [WebCite Cache ID 6tEcfAKsx]
27. Mohebbi M, Vanderkam D, Kodysh J, Schonberger R, Choi H, Kumar S. Google. Google correlate whitepaper URL: https://www.google.com/trends/correlate/whitepaper.pdf [accessed 2017-09-04] [WebCite Cache ID 6tEca1Xnd]
28. Zhang W. 2013. Development of a real-time estimate of flu activity in the United States using dynamically updated lasso regressions and Google search queries URL: http://www.people.fas.harvard.edu/~msantill/Mauricio_Santillana/Teaching_files/D_Zhang_thesis_final.pdf [accessed 2017-09-04] [WebCite Cache ID 6tEdMtiVP]
29. Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J. Freebase: a collaboratively created graph database for structuring human knowledge. 2008 Presented at: ACM SIGMOD International Conference on Management of Data; June 9-12, 2008; Vancouver, BC, Canada.

30. Durbin J, Koopman S. Time series analysis by state space methods. Oxford, UK: Oxford University Press; 2012.
31. Hamilton JD. Time Series Analysis. Princeton, NJ: Princeton University Press; 1994.
32. Ripley BD. 2002. Time series in R 1.5.0 URL: https://www.r-project.org/doc/Rnews/Rnews_2002-2.pdf [accessed 2017-10-08] [WebCite Cache ID 6u3R57laM]
33. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
34. Broniatowski DA, Dredze M, Paul MJ, Dugas A. Using social media to perform local influenza surveillance in an inner-city hospital: a retrospective observational study. JMIR Public Health Surveill 2015;1(1):e5 [FREE Full text] [doi: 10.2196/publichealth.4472] [Medline: 27014744]
35. Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. J Stat Softw 2008;27(3):2008. [doi: 10.18637/jss.v027.i03]
36. Hyndman R. Forecasting functions for time series and linear models URL: https://cran.r-project.org/web/packages/forecast/index.html [accessed 2017-09-04] [WebCite Cache ID 6tEbWlup4]
37. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer; 2009.
38. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
39. Breiman L. 2002. Manual on setting up, using, and understanding Random Forests v3.1 URL: https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf [accessed 2017-09-04] [WebCite Cache ID 6tEdmjBnR]
40. CDC. MMWR weeks URL: https://wwwn.cdc.gov/nndss/document/MMWR_week_overview.pdf [accessed 2017-09-04] [WebCite Cache ID 6tEdRkN5n]
41. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc 1937 Dec;32(200):675-701. [doi: 10.1080/01621459.1937.10503522]
42. Hollander M, Wolfe DA, Chicken E. Nonparametric Statistical Methods. Hoboken, NJ: Wiley; 2013.
43. Nemenyi P. Distribution-free Multiple Comparisons. Princeton, NJ: Princeton University; 1963.
44. Pohlert T. 2014. PMCMR: calculate pairwise multiple comparisons of mean rank sums URL: https://cran.r-project.org/web/packages/PMCMR/index.html [accessed 2017-09-04] [WebCite Cache ID 6tEbMqPK0]
45. Liaw A, Wiener M. Cogns.northwestern. 2002. Classification and regression by randomForest URL: http://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf [WebCite Cache ID 6u3SN9cRB]
46. R core team. R-project. 2013. R: A language and environment for statistical computing URL: http://www.r-project.org/ [WebCite Cache ID 6tEdZQFnY]

Abbreviations

API: application programming interface
ARIMA: autoregressive integrated moving average
CDC: Centers for Disease Control and Prevention
GET: Google Extended Trends
GFT: Google Flu Trends
HHS: US Department of Health and Human Services
ILI: influenza-like illness
ILINet: US Outpatient Influenza-like Illness Surveillance Network
IQR: interquartile range
MAPE: mean absolute percentage error
MMWR: Morbidity and Mortality Weekly Report
RMSE: root mean square error

Edited by A Keepanasseril; submitted 10.02.17; peer-reviewed by M Santillana, D Broniatowski; comments to author 07.04.17; revised version received 13.06.17; accepted 15.08.17; published 06.11.17.

Please cite as:
Kandula S, Hsu D, Shaman J
Subregional Nowcasts of Seasonal Influenza Using Search Trends
J Med Internet Res 2017;19(11):e370
URL: http://www.jmir.org/2017/11/e370/
doi:10.2196/jmir.7486
PMID:29109069

Original Paper

Understanding Older Adult's Technology Adoption and Withdrawal for Elderly Care and Education: Mixed Method Analysis from National Survey

Ching-Ju Chiu[1], PhD; Chia-Wen Liu[2], BS

[1]Institute of Gerontology, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[2]Institute of Physical Therapy, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Corresponding Author:
Ching-Ju Chiu, PhD
Institute of Gerontology, College of Medicine
National Cheng Kung University
No 1, University Road
Tainan, 70101
Taiwan
Phone: 886 62353535 ext 5739
Fax: 886 63028175
Email: cjchiu@mail.ncku.edu.tw

Abstract

Background: Elderly adults have comprised the fastest growing population adopting the Internet and computer technology over the past decade. However, how their experiences can shed light on elderly learning theory has not been examined much in the literature.

Objective: This study investigated the factors and reasons associated with Internet adoption and withdrawal among older adults in Taiwan, and if any gender differences exist in this context.

Methods: Data on participants aged 50 years and older from the nationally representative "Digital Opportunity Survey on Individuals and Households in Taiwan," who did not use the Internet in 2005 but adopted it in 2007 (n=1548), and those who reported using Internet in 2011 but then withdrew (n=1575), were analyzed. Factors and reasons associated with Internet adoption and withdrawal were examined using both quantitative and qualitative data.

Results: Education level independently predicted Internet adoption behavior. With regard to the reasons for adoption, 66% (62/94) of participants indicated they started using the Internet to meet certain "needs"; for example, "keeping up with the world" (40.4%, 38/94) was listed as the most critical reason, followed by "job needs" (25.5%, 24/94). Older adults with a positive attitude toward the Internet with regard to increasing employment opportunities (OR 2.0, 95% CI 1.0-3.9, *P*=.04) and the amount of information obtained (OR 0.5, 95% CI 0.3-0.9, *P*=.01), as well as enriching recreation and entertainment (OR 0.6, 95% CI 0.4-0.9, *P*=.02), were less likely to withdraw from the Internet. The most common reason for Internet withdrawal was "psychological barriers" (eg, no available time, no meaningful use, or nothing worth reading/watching; 66.3%, 193/291), followed by "health barriers" (eg, eyes or body deteriorate with Internet use; 21.0%, 61/291). Although psychological barriers were the most important factor for Internet withdrawal for both men (72.5%, 100/138) and women (62%, 93/150), women were more likely than men to be affected by health barriers (26.0%, 39/150 vs 15.9%, 22/138; *P*=.004) and anthropic factors or accidental barriers (7.3%, 11/150 vs 2.9%, 4/138; *P*=.02).

Conclusions: Our findings that the need to keep up with the world associated with Internet adoption, and gender differences in reasons behind Internet withdrawal, such that women reported more health and anthropic factors or accidental barriers than man, may provide a new perspective that help health educators understand strategies that encourage older adults to keep learning, an important component of active aging.

(J Med Internet Res 2017;19(11):e374) doi:10.2196/jmir.7401

KEYWORDS

Internet adoption; Internet withdrawal; digital opportunity; information literacy; middle-aged and older adults; Taiwan

Introduction

Demographic statistics indicate that 8.1% of the global population were older adults in 1960, and this number grew to 10% in 2000, with estimates suggesting that 21.4% of the population will be senior citizens in 2050 [1]. A demographic survey in the United States indicated that the elderly population will increase from 13% in 2010 to 19% in 2030 [2]. Internet use is increasingly widespread, although the number of people who do not use the Internet frequently, such as the elderly, is in fact increasing [3]. Despite this, the first nationally representative study in the United States showed that the proportion of seniors (those aged 65 years and older; mean 75, SD 7.4 years) who use any digital health technologies significantly increased from 21% in 2011 to 25% in 2014 [4].

Internet use can bring many benefits, such as increasing the happiness and decreasing the loneliness of older adults in a retirement community [5], encouraging social connections, and accelerating information exchanges among adults older than 50 years in the United States [6]. Studies have shown that the social functions that come from Internet use contribute to the maintenance of relationships for older adults aged 57 to 87 years [7], particularly for older people with limited mobility, because it is a good way for them to come out of solitude and reconnect to the rest of the world [6]. One of the significant aims of active aging is to keep learning in order to adapt to the changes that occur in later life, gain capabilities similar to those of younger people, and remain productive [8], and it is believed that older adults' involvement in the Internet may achieve this [9-13]. But the Internet is associated with security concerns and learning anxiety (eg, how to handle the operating systems, how to resolve any problems encountered, and the fear that one might break an expensive device) [14,15]. Not every population group benefits from Internet use, and there can even be certain negative impacts. However, many studies in the United States indicate that there are more positive than negative impacts arising from Internet use [16-18].

Many studies have been conducted on why older people may choose to avoid the Internet, such as the webpage design not being suitable for older people and the perception of not being able to learn new things because of insufficient cognitive capability, vision, or motor function [19-27]. The diverse needs of older users should be considered in the design and development of such technology [13]. However, there are still only a handful of studies that aim to understand Internet withdrawal among middle-aged and older adults. A previous study showed that the objective of e-learning was not only to describe something, but also to demonstrate how to do it [28]. Knowles' adult education theory [29] highlights the importance of self-direction, a reservoir of experience, social roles, and problem-centered orientation in affecting adult learning, and Internet adoption and withdrawal may be affected by these factors. For example, the Internet creates an independent learning opportunity for the self-directed learning of adults, while family or friends can serve as a facilitator. Older learners have a large body of experience to serve as a background for new learning, and they are strongly motivated to learn information that has immediate application [30]. In contrast, if adults feel they are unable to learn, they are more likely to withdraw from Internet use.

Gender differences may also exist in Internet adoption and withdrawal among middle-aged and older adults. Previous studies indicated that website attributes [31], Internet usage patterns [32], and habits [33] are different across genders. More males use the Internet than females [32], and the differences in their visual cues have a great impact on their online choices [34]. Women have also expressed greater levels of anxiety toward computers [35], less self-perceived competence, and lower perceived ease of use with respect to the Internet [36] than men. Moreover, older males seem to perceive the Internet as more useful due to their perceived higher levels of ease of use than females [37]. However, some studies found no differences across genders in these respects [38,39]. It is thus an interesting issue whether there are any gender differences in Internet adoption and withdrawal. If we could underline different characteristics associated with gender, it would encourage the development of genuinely usable information and communications technology products, training, and support approaches, and it could be used as a basis for the design of continuing education materials or for market segmentation of men and women.

In Taiwan, there is little data on older adults' Internet use, with no nationwide surveys. As a result, there is no solid reference material for Internet adoption and withdrawal among middle-aged and older adults in Taiwan. In addition, because older adults are the fastest growing population in adopting the Internet and computer technology during the past decade [40], depicting factors and reasons associated with Internet adoption and withdrawal may shed light on some aspects of elderly learning theory. Thus, the purposes of this study were (1) to understand the prevalence of Internet adoption and withdrawal among middle-aged and older adults in Taiwan, (2) to examine the factors associated with Internet adoption and withdrawal among middle-aged and older adults in Taiwan, (3) to identify the reasons behind those who had Internet adoption and withdrawal, and (4) investigate if the aforementioned patterns differed by gender.

Methods

Study Participants and Data Sources

Data were derived from an on-going survey, the Individual & Household Digital Opportunity Survey, approved by the National Development Council in its Research on Constructing the Index System of Digital Opportunity Development in Taiwan. It covers home phone users and interviews native Taiwan citizens aged 12 years or older residing in ordinary households, and it has been carried out every year since 2002. Computer-assisted telephone interviews were adopted in this survey using a random stratified sampling procedure.

Each year's survey mainly covers information on Internet access, information literacy, information application, digital opportunity, and digital exclusion. Each year's survey is designed as a cross-sectional study of that year. However, only one follow-up survey was conducted, with the first and second surveys taking

place in 2005 and 2007. Those participants who did not use the Internet in 2005 were asked again in 2007 whether they had done so. As a result, we used the data in 2007 for the issue of Internet adoption. In addition, only the 2011 survey asked a question about Internet withdrawal; therefore, the 2011 survey was selected to examine the issue of Internet withdrawal. A total of 1548 valid participants aged 50 years and older in 2007 (success rate: 72.4%) and 1575 adults in 2011 (success rate: 69.4%) were analyzed in this study.

Measures

Sociodemographic Variables

Age (range 51-94 years), gender, living area (northern, central, southern, and eastern Taiwan and outer islands), economic status (with an income of Taiwanese new dollar [NT$] NT$30,000 or less, between NT$30,000 and NT$90,000, and NT$90,000 or more per month), and educational background (elementary school or lower, high school, college or higher) were recorded.

Internet Adoption/Withdrawal

In this study, those who did not use the Internet in 2005 were asked in 2007 whether they now did. Those who answered "yes" were defined as Internet adopters, whereas those who answered "no" were defined as Internet nonadopters. For Internet withdrawal, the definition was based on the Internet use experience in 2011 and whether the person had used the Internet in the last month. Those who had previous experience of Internet use and had used it in the last month were defined as Internet nonwithdrawers, whereas those who had previous experience of Internet use but had not used it for a month were defined as Internet withdrawers. The reasons for Internet adoption and withdrawal were obtained from an open-ended question. Participants were asked: "Why did you start to use the Internet?" in 2007 and "Why don't you use the Internet anymore?" in 2011.

Opinions on Internet Use

Opinions on Internet use were divided into three dimensions of "perceived helpfulness," "perceived fun," and "perceived interpersonal interaction." The question for perceived helpfulness was "Is Internet use in your opinion helpful to your life?" and the choices were 1=very helpful, 2=somewhat helpful, and 3=not helpful. The question for perceived fun was "Do you feel that Internet use gives you more fun, less fun, or no effect at all?" and the choices were 1=less fun, 2=no effect, and 3=more fun. The question for perceived interpersonal interaction was "Does the Internet allow you to interact with your friends and relatives more frequently, less frequently, or have no effect at all?" and the choices were 1=more frequently, 2=less frequently, and 3=no effect.

Digital Opportunity

"Digital opportunity" referred to whether the respondents felt any change or opportunities created from starting to use the Internet in their daily lives (eg, opportunities for employment and learning, or changes in their circle of friends or income). The study participants were asked if they had noticed any of the following changes: "Has your circle of friends / employment opportunity / learning opportunity / income / access to useful information / recreation and entertainment / government information increased because of your Internet use?" and "Is it possible for you to connect with others who share the same views in politics or policies?" (eight questions in total). The choices were 1=yes and 2=no.

Information Literacy

"Information literacy" referred to the basic capabilities that are required for Internet use (eg, becoming a member of specific website; downloading and uploading files, video clips, or photos). The study participants were asked, "Have you applied for user accounts and passwords and become a member of specific website?" and "Do you know how to download and upload files, video clips, or photos?" The choices were 1=yes and 2=no. They were also asked, "Are you familiar with any kind of word processing program (eg, Word, Notepad, or Writer) for document editing?" and the choices were 1=very familiar, 2=somewhat familiar, 3=not very familiar, and 4=no idea whatsoever. The distribution of information literacy was determined based on these categories.

Statistical Analysis

To examine the sociodemographic characteristics, opinions for Internet use, digital opportunity, and information literacy between participants with and without Internet adoption or withdrawal, *t* tests and chi-square tests were used for the ordinal and nominal variables, respectively. Second, logistic regression analysis was employed to examine independent factors predicting Internet adoption and withdrawal with comparison to those individuals without adoption or withdrawal. Third, open data coding was carried out for the reasons for Internet adoption and withdrawal, with the results described using the frequency and percentage. We finally examined gender differences with regard to the aforementioned patterns, stratifying the analysis by gender.

Results

Sociodemographic Characteristics and Opinions on Internet Use, Digital Opportunity, and Information Literacy

The distribution of demographic characteristics and opinions about Internet use, digital opportunity, and information literacy of the participants with and without Internet adoption or withdrawal is presented in Table 1. Among the 1548 participants in 2007, only 6.65% (n=103) adults belonged to the Internet adoption group (nonadopters: 1445/1548, 93.35%). Among the 1575 adults in 2011, only 18.60% (n=293) belonged to the Internet withdrawal group (nonwithdrawers: 1282/1575, 81.40%). When it came to adopting the Internet, the factors of age, educational background, family income per month, and influence on life were significantly different between the adopters and nonadopters, whereas gender was not. Most of the Internet adopters had a high school or above diploma (89/103, 86.4%); nearly 60% (60/103) of the adopters had a family income of more than NT$30,000 per month. Most of the adopters believed the Internet helped with their lives (85/103, 85.9%), and 60.2% (62/103) felt that it introduced more fun in their lives.

Table 1. Sociodemographic characteristics, opinions on Internet use, digital opportunity, and information literacy between participants who did and did not adopt the Internet or withdraw from the Internet.

Item by Internet adoption or withdrawal	Yes (adopted/withdrew)	No (did not adopt/withdraw)	*P*
Internet adoption	**N=103**	**N=1445**	
Age (years), mean (SD)	58.4 (6.0)	65.2 (9.6)	<.001
Gender (male), n (%)	54 (52.4)	681 (47.1)	.30
Education, n (%)			<.001
Elementary school or below	14 (13.6)	1043 (72.2)	
High school	66 (64.1)	333 (23.1)	
College or above	23 (22.3)	69 (4.8)	
Family income/month (NT$), n (%)			<.001
<30,000	43 (41.3)	986 (68.2)	
30,000~90,000	45 (44.0)	384 (26.6)	
>90,000	15 (14.7)	75 (5.2)	
Opinions on Internet use, n (%)			
Perceived helpfulness	85 (85.9)	—	—
Perceived fun	62 (60.2)	—	—
Perceived more interpersonal interaction	21 (20.4)	—	—
Internet withdrawal	**N=293**	**N=1282**	
Age (years), mean (SD)	59.6 (7.1)	57.6 (5.6)	<.001
Gender (male), n (%)	139 (47.4)	705 (55.0)	.02
Education, n (%)			<.001
Elementary school or below	40 (13.7)	58 (4.5)	
High school	173 (58.9)	524 (40.9)	
College or above	80 (27.4)	700 (54.6)	
Family income/month (NT$), n (%)			<.001
<30,000	82 (27.9)	146 (11.4)	
30,000~90,000	152 (52.1)	563 (43.9)	
>90,000	59 (20.0)	573 (44.7)	
Digital opportunity, n (%)			
Larger circle of friends	72 (24.6)	459 (35.8)	.001
More job opportunities	35 (12.0)	237 (18.5)	.009
More learning opportunities	123 (41.8)	948 (73.9)	<.001
Increased income	16 (5.5)	142 (11.1)	.004
More information of life	120 (41.0)	997 (77.8)	<.001
Enriched recreation and entertainment	114 (39.0)	914 (71.3)	<.001
More government information	90 (30.7)	761 (59.4)	<.001
Ability to connect with others with the same political/policy views	14 (4.6)	213 (16.6)	<.001
Information literacy, n (%)			
Know how to apply for a user account and password	52 (17.8)	705 (55.0)	<.001
Know how to use a word processing program	135 (46.1)	885 (69.0)	<.001
Know how to upload and download files	99 (33.8)	883 (68.9)	<.001

The analysis of Internet withdrawal shows that the factors of age, gender, educational background, family income per month, digital opportunity, and information literacy were significantly different between those who withdrew or did not, whereas where they lived did not. For the withdrawers, most were female (154/293, 52.6%) and 72.6% of them (213/293) had a high school or lower diploma. Approximately 80% of the withdrawers (234/293) had a family income per month of NT$ 90,000 or less. The top three digital opportunities that the nonwithdrawers had were Internet use helps with the "improved learning opportunity" (123/293, 41.8%), followed by "more access to useful information" (120/293, 41.0%), then "enriched recreation and entertainment" (114/293, 39.0%). In addition, less than 50% of the withdrawers (135/293, 46.1%) selected "know how to use a word processing program (eg, Word or Notepad) for document editing" for their information literacy, whereas the percentage was greater than 50% for nonwithdrawers in all three categories of information literacy.

Factors Predicting Internet Adoption and Withdrawal

The results of the logistic regression analysis evaluating the independent effects of the demographic characteristics, opinions on Internet use, digital opportunity, and information literacy of the study participants on Internet adoption and withdrawal are presented in Table 2. For Internet adoption, the older the participants were, the less likely they were to experience adoption (OR 0.9, 95% CI 0.92-0.9), and similar results for educational background, as those with less education had less adoption. For withdrawal, the Internet withdrawal risk was significantly lower for middle-aged and older adults who had a family income per month of NT$90,000 or more (as opposed to those with NT$30,000 or less; OR 0.4, 95% CI 0.2-0.8), who felt enriched recreation and entertainment (OR 0.6, 95% CI 0.4-0.9), access to more useful information (OR 0.5, 95% CI 0.3-0.9), knew how to apply for a user account and password (OR 0.4, 95% CI 0.2-0.6), and how use a word processor (OR 0.6, 95% CI 0.4-0.9). On the other hand, the risk of Internet withdrawal was significantly higher for those participants who were older (OR 1.1, 95% CI 1.0-1.1).

Table 2. Factors predicting Internet adoption and withdrawal among middle-aged and older adults by logistic regression.

Factor	Adoption (yes/no)		Withdrawal (yes/no)	
	OR (95% CI)	*P*	OR (95% CI)	*P*
Age	0.9 (0.9-0.9)	.001	1.1 (1.0-1.1)	.003
Gender (male/female)	0.8 (0.4-1.3)	.28	0.7 (0.4-1.0)	.07
Educational background				
High/elementary school or below	10.0 (4.9-20.4)	<.001	0.9 (0.4-2.0)	.89
College or above/elementary school or below	19.4 (7.9-47.1)	<.001	0.5 (0.2-1.1)	.10
Family income/month (NT$)				
30,000-90,000/<30,000	0.9 (0.5-1.7)	.85	0.8 (0.5-1.4)	.39
>90,000/<30,000	1.0 (0.4-2.4)	.10	0.4 (0.2-0.8)	.008
Digital opportunity (yes/no)				
larger circle of friends	—	—	1.4 (0.9-2.3)	.19
More job opportunities	—	—	2.0 (1.0-3.9)	.04
More learning opportunities	—	—	0.8 (0.5-1.3)	.31
Increased income	—	—	0.5 (0.2-1.4)	.17
More life information	—	—	0.5 (0.3-0.9)	.01
Enriched recreation and entertainment	—	—	0.6 (0.4-0.9)	.02
More government information	—	—	0.8 (0.5-1.2)	.28
Ability to connect with others who have the same political/policy views	—	—	0.6 (0.3-1.4)	.24
Information literacy (yes/no)				
Know how to apply for a user account and password	—	—	0.4 (0.2-0.6)	<.001
Know how to use a word processing program	—	—	1.2 (0.8-1.9)	.43
Know how to upload and download files	—	—	0.6 (0.4-0.9)	.03

Table 3. Reasons for Internet adoption and withdrawal for middle-aged and older adults.

Type and item	n (%)
Internet adoption	**N=94**
Needs (66.0%)	
Job needs	24 (25.5)
Keep up with the world	38 (40.4)
Expansion of life (24.4%)	
Look for data and information	9 (9.6)
Stock market	6 (6.4)
Interested in learning	4 (4.3)
Online ticket purchase	2 (2.1)
Read news	2 (2.1)
Recreation and entertainment (9.6%)	
Play games and have fun	4 (4.3)
Kill some time	4 (4.3)
Watch TV online	1 (1.1)
Internet withdrawal [a]	**N=291**
Health barriers (21.0%)	
I am old and my eyes or body has deteriorated	48 (16.5)
I forgot how it works; it does not work well for me	13 (4.5)
Psychological barriers (66.3%)	
I do not have the time	128 (44.0)
It has no use for me	53 (18.2)
Nothing worth watching/reading	12 (4.1)
Equipment or environmental barriers (7.2%)	
The computer is out of order	10 (3.4)
I do not have a computer or Internet access at home	8 (2.8)
Internet access is too expensive	1 (0.3)
Internet connection is poor	1 (0.3)
Stop for home renovation	1 (0.3)
Anthropic factor or accidental barriers (5.5%)	
Others need the computer at home	8 (2.8)
Travel to a foreign country or stay somewhere else for a period of time	3 (1.0)
Family accident	2 (0.7)
Internet access is canceled so that kids will not use it too much	2 (0.7)
Retirement	1 (0.3)

[a]The sample size of Internet withdrawers was 293 adults and there were two with missing data; therefore, the final sample size of Internet withdrawers was 291 older adults.

Reasons for Internet Adoption and Withdrawal

Table 3 shows the results of the causes for middle-aged and older Internet adopters and withdrawers. For the reason of adoption, 66% (62/94) of participants indicated they started using Internet "out of need," followed by "expansion of life" (23/94, 24.4%), and then "recreation and entertainment" (9/94, 9.6%). Among "needs," "keep up with the world" (38/94, 40.4%) was listed as the most critical reason, followed by "job needs" (24/94, 25.5%). For the reason of withdrawal, the highest percentage went to "psychological barriers," such as having no time available, no meaningful use, or nothing worth reading/watching (193/291, 66.3%); followed by "health barriers" such as eyes or body deteriorating or do not work well enough to use the Internet (61/291, 21.0%), "equipment barriers" such as having no computer or Internet access at home (21/291,

7.2%), and "anthropic factor or accidental barriers" such as having no computer to use because others need it at home, or travel to a foreign country for a period of time (16/291, 5.5%).

Gender Differences

The results of the analysis for the reasons for Internet withdrawal among the middle-aged and older adults by gender are shown in Table 4. For the details of withdrawal reason, a large proportion of Internet withdrawals could be attributed to having no time available (males: 67/138, 44.7%; females: 61/150, 44.2%). However, for women the second most significant reason for withdrawal was "I am old and my eyes or body has deteriorated" (32/150, 21.3%) and the third was "it has no use for me" (21/150, 14.0%). For men, the second most significant reason for withdrawal was "it has no use for me" (32/138, 23.2%) and the third was "I am old and my eyes or body has deteriorated" (16/138, 11.6%). For the category of withdrawal reason, the motivation for withdrawal was significantly more often due to anthropic factor or accidental barriers (11/150, 7.3% vs 4/138, 2.9%, *P*=.02) or health barriers (39/150, 26.0% vs 22/138, 16.0%, *P*=.004) for women than for men. Moreover, despite the failure to reach statistical significance, men had more psychological and equipment-related barriers than women. Nevertheless, it was found that the psychological barriers were the most important factor influencing Internet withdrawal for both males and females. A total of 72.5% (100/138) males and 62.0% (93/150) females believed that their Internet withdrawal behaviors were associated with psychological factors such as having no time available, no meaningful use, or nothing worth watching/reading.

Table 4. Reasons for Internet withdrawal among middle-aged and older men (n=138) and women (n=150).

Reason for Internet withdrawal	Men, n (%)	Women, n (%)	*P*
Reasons (by "details")			
Health barriers			
I am old and my eyes or body has deteriorated	16 (11.6)	32 (21.3)	.003
I forgot how it works; it does not work well for me	6 (4.4)	7 (4.7)	.52
Psychological barriers			
I do not have the time	61 (44.2)	67 (44.7)	.11
It has no use for me	32 (23.2)	21 (14.0)	.44
Nothing worth watching/reading	7 (5.1)	5 (3.3)	.82
Equipment or environmental barriers			
The computer is out of order	6 (4.4)	4 (2.7)	.75
I do not have a computer or Internet access at home	5 (3.6)	3 (2.0)	.67
Internet access is too expensive	1 (0.7)	0 (0.0)	.37
Internet connection is poor	0 (0.0)	1 (0.7)	.27
Stop for home renovation	1 (0.7)	0 (0.0)	.37
Anthropic factor or accidental barriers			
Others need the computer at home	1 (0.7)	7 (4.7)	.02
Travel to a foreign country or stay at somewhere else for a period of time	0 (0.0)	3 (2.0)	.06
Family accident	2 (1.5)	0 (0.0)	.20
Reason (by "category")			
Health barriers	22 (16.0)	39 (26.0)	.004
Psychological barriers	100 (72.5)	93 (62.0)	.41
Equipment or environmental barriers	12 (8.7)	7 (4.7)	.53
Anthropic factor or accidental barriers	4 (2.9)	11 (7.3)	.02

Discussion

This study is the first to use nationally representative data for older adults in Taiwan to examine the issues of Internet adoption and withdrawal. The results showed that both the respondents who adopted the Internet and with greater probability of continued use had the characteristics of higher proportion in men, younger and with higher education. In addition, the most important reasons for Internet adoption were associated with "needs," especially keeping up with the world and job needs. Furthermore, although psychological barriers were the most important factor of Internet withdrawal for both men and women, women were more likely to be affected by health and anthropic factors or accidental barriers in this regard.

A number of factors associated with older adults' technology adoption have been documented in the literature [41,42], although the reasons behind these factors have rarely been

examined. This study found that the main reasons why older adults may choose to use the Internet are associated with needs, especially with regard to keeping up with the world and job-related needs. In addition, if older adults think that using the Internet can increase their employment opportunities (OR 2.0, 95% CI 1.0-3.9, *P*=.04), access to useful information (OR 0.5, 95% CI 0.3-0.9, *P*=.01), and recreation and entertainment, then they are less likely to withdraw from using it. These findings echo previous research indicating that older adults who do not use new technology or learn new things may simply not see the need for much of what is being offered [43]. Education or training should be provided to help older adults understand the underlying structures and benefits of new learning opportunities.

Many of the existing studies focused on how deteriorating health conditions can hinder older adults' use of technology or learning efforts. Recent international studies also indicated that people tend to withdraw from the Internet due to health-related factors (eg, age-related changes in visual acuity, color perception and susceptibility to glare, and hearing problems) and some psychological barriers (eg, computer anxiety, online problems, and privacy issues) [15,16,44-46]. This study used a nationally representative sample and the results echo those of other recent works, showing that psychological barriers play the most important role in Internet withdrawal among middle-aged and older adults in Taiwan, followed by health factors. Moreover, it was surprising to find that most of the middle-aged and older adults in this study's data felt that "no available time" was the main reason that kept them from not using Internet. This suggests that elderly education should not only focus on health conditions because psychological factors may be critical to Internet usage and learning. For example, the selective optimization with compensation model of successful aging [47] may be useful in helping older adults to cope with the barriers they face when adopting new learning. Compensation reminds older adults to consider the reality of a person's capacities and the health barriers they face, selection refers to the ways older adults use to overcome such barriers, whereas optimization highlights the resources older adults have for achieving their goals.

Gender differences in Internet use have been documented in a number of studies, although in this study using Taiwanese data, men and women were not different in Internet adoption. However, they were different in Internet withdrawal, with men less likely to withdraw from Internet use than women were. In addition, for both men and women, the most common response for the reason for Internet withdrawal was for "psychological barriers" (eg, no time available, no meaningful use, or nothing worth reading/watching) followed by "health barriers" (eg, eyes or body deteriorating with Internet use). Although "psychological barriers" were the most important factor for Internet withdrawal among both men and women, women were more likely than men were to be affected by health and anthropic factors or accidental barriers. We suggest that in addition to focusing on health barriers, such as relearning to overcome cognitive declines, it is also important to work to reduce anthropic factors or accidental barriers, possibly by providing more resources, to encourage women to use new technology or take more learning opportunities.

This study has some limitations. First, this work was based on a cross-sectional survey and it may not be possible to draw any conclusions on the causal relationship between Internet user types and social engagement. Second, due to our data structure we only investigated Internet adoption and withdrawal behavior during 2007 and 2011. Given the rapidly changing nature of the Internet and technology adoption by older adults, we acknowledge that the dynamics of adoption and withdrawal may change over time. However, although the next generation of older adults will have extensive Internet experience, different problems and solutions may arise with regard to adoption and withdrawal behaviors due to the nature of the human learning process. Moreover, the rapid pace of technological change means that "future older generations" are likely to confront an array of technologies they little understand and generally find inaccessible. According to Hanson [48], understanding the general technology-related skills of older users, identifying the strategies successfully used by this population, and finding designs that are optimized for older adults' abilities (eg, life experiences and knowledge) are the most promising directions for research into technology or computer use by older adults. We believe that factors and reasons associated with Internet adoption and withdrawal found in this study can help to inform education or care for the elderly, a subpopulation that is not in the mainstream or main working force of a nation, and who have special needs associated with their physical and psychological degeneration. Third, due to the limitations of secondary data, this study only examined basic sociodemographic correlates. Various factors that may be related to Internet adoption or withdrawal, such as having a disability or living alone, should be examined in future research. Fourth, the definition of Internet withdrawal remains unclear in the literature. In this study, we defined it as if those participants with Internet experience had not used the Internet during the previous month. This definition is justified because, based on the cognitive competence and memory function of older adults, if such individuals do not use the Internet for one month, then their likelihood of doing so again decreases. However, the specific definition of Internet withdrawal among older adults used in the literature needs to be explored more in the future.

A critical role of technology as an important health promotion strategy for older adults in low- and middle-income countries has been proposed [49]. Our study on factors associated with Internet adoption and withdrawal provides a new lens that can help health educators to understand strategies that foster older adults in learning, an important element for active aging. Specifically, this study examining factors and reasons associated with Internet adoption and withdrawal from nationally representative data on middle-aged and older adults in Taiwan found that the learning motivation and learning models for older adults may not be different by gender, but to prevent withdrawal from learning by middle-aged and older adults, it is important that the approach be different by gender. With our findings that women were more likely to be affected by health and anthropic factors or accidental barriers, we suggest that in addition to focusing on health barriers, such as relearning to overcome

cognitive declines, it is also important to work to reduce anthropic factors or accidental barriers, possibly by providing more resources, to encourage women to take more learning opportunities. In addition, for health educators to design courses and activities, it is essential that learning satisfies older adults' needs, such as keeping up with the world or job needs, such that they gain the knowledge and skills that could increase quality of life and assist them to transition into aging successfully.

Acknowledgments

This work was supported by grants from the Ministry of Science and Technology in Taiwan (MST103-2314-B-006-038-MY3).

Conflicts of Interest

None declared.

References

1. Cohen JE. Human population: the next half century. Science 2003 Nov 14;302(5648):1172-1175. [doi: 10.1126/science.1088665] [Medline: 14615528]
2. Vincent G, Velkoff V. The Next Four Decades. The Older Population in the United States: 2010 to 2050. Washington, DC: US Department of Commerce, Economics and Statistics Administration, US Census Bureau; 2010. URL: https://www.census.gov/prod/2010pubs/p25-1138.pdf [accessed 2017-10-30] [WebCite Cache ID 6ubUnaZXQ]
3. Fox S. Pew Research Center. 2011. Health topics URL: http://www.pewinternet.org/2011/02/01/health-topics-2/ [accessed 2017-10-23] [WebCite Cache ID 6uQLOHerY]
4. Levine DM, Lipsitz SR, Linder JA. Trends in seniors' use of digital health technology in the United States, 2011-2014. JAMA 2016 Aug 2;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]
5. Shapira N, Barak A, Gal I. Promoting older adults' well-being through Internet training and use. Aging Ment Health 2007 Sep;11(5):477-484. [doi: 10.1080/13607860601086546] [Medline: 17882585]
6. Hogeboom DL, McDermott RJ, Perrin KM, Osman H, Bell-Ellison BA. Internet use and social networking among middle aged and older adults. Educ Gerontol 2010 Jan 05;36(2):93-111. [doi: 10.1080/03601270903058507]
7. Trocchia PJ, Janda S. A phenomenological investigation of Internet usage among older individuals. J Consum Mark 2000 Dec;17(7):605-616. [doi: 10.1108/07363760010357804]
8. Torres S. A culturally-relevant theoretical framework for the study of successful ageing. Ageing Soc 1999;19(1):33-51.
9. Nimrod G. The benefits of and constraints to participation in seniors' online communities. Leisure Stud 2012 Jul 12;33(3):247-266. [doi: 10.1080/02614367.2012.697697]
10. Berner J, Aartsen M, Wahlberg M, Elmståhl S, Berglund J, Anderberg P, et al. A cross-national and longitudinal study on predictors in starting and stopping Internet use (2001-2013) by Swedish and Dutch older adults 66 years and above. Gerontechnology 2016;14(3):157-168. [doi: 10.4017/gt.2016.14.3.012.00]
11. Lifshitz R, Nimrod G, Bachner YG. Internet use and well-being in later life: a functional approach. Aging Ment Health 2016 Sep 22:1-7. [doi: 10.1080/13607863.2016.1232370] [Medline: 27657190]
12. Friemel TN. The digital divide has grown old: determinants of a digital divide among seniors. New Media Soc 2014 Jun 12;18(2):313-331. [doi: 10.1177/1461444814538648]
13. Dove E, Astell A. The use of motion-based technology for people living with dementia or mild cognitive impairment: a literature review. J Med Internet Res 2017;19(1):e3. [doi: 10.2196/jmir.6518]
14. Cody MJ, Dunn D, Hoppin S, Wendt P. Silver surfers: training and evaluating internet use among older adult learners. Commun Educ 1999 Oct;48(4):269-286. [doi: 10.1080/03634529909379178]
15. Carpenter BD, Buday S. Computer use among older adults in a naturally occurring retirement community. Comput Hum Behav 2007 Nov;23(6):3012-3024. [doi: 10.1016/j.chb.2006.08.015]
16. Gatto SL, Tak SH. Computer, Internet, and e-mail use among older adults: benefits and barriers. Educ Gerontol 2008 Aug 22;34(9):800-811. [doi: 10.1080/03601270802243697]
17. Jelenchick LA, Eickhoff JC, Moreno MA. Facebook depression? Social networking site use and depression in older adolescents. J Adolesc Health 2013 Jan;52(1):128-130. [doi: 10.1016/j.jadohealth.2012.05.008] [Medline: 23260846]
18. Taha J, Sharit J, Czaja S. Use of and satisfaction with sources of health information among older Internet users and nonusers. Gerontologist 2009 Oct;49(5):663-673. [doi: 10.1093/geront/gnp058] [Medline: 19741112]
19. Oppenauer C. Motivation and needs for technology use in old age. Gerontechnology 2009;8(2):82. [doi: 10.4017/gt.2009.08.02.006.00]
20. Burnett JS, Mitzner TL, Charness N, Rogers WA. Understanding predictors of computer communication technology use by older adults. Hum Fac Erg Soc P 2011 Sep 01;55(1):172-176. [doi: 10.1177/1071181311551036]
21. Von Solms R, De Lange E. SenPortal: Internet-based services for senior citizens: case notes: building the information society. Afr J Info Comm 2013:76-81.

22. Yu C. Computer and Internet Attitudes Among Ethnically Diverse Older Adults: A Path Analysis [master's thesis]. Northridge, CA: California State University, Northridge; 2012.
23. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
24. Werner JM, Carlson M, Jordan-Marsh M, Clark F. Predictors of computer use in community-dwelling, ethnically diverse older adults. Hum Factors 2011 Oct;53(5):431-447 [FREE Full text] [Medline: 22046718]
25. Kim KO. The Emotional Responses of Older Adults to New Technology [dissertation]. Urbana, IL: University of Illinois at Urbana-Champaign; 2012.
26. Adams N, Stubbs D, Woods V. Psychological barriers to Internet usage among older adults in the UK. Med Inform Internet Med 2005 Mar;30(1):3-17. [doi: 10.1080/14639230500066876] [Medline: 16036626]
27. Berner JS, Rennemark M, Jogréus C, Berglund J. Factors associated with change in Internet usage of Swedish older adults (2004-2010). Health Informatics J 2013 Jun;19(2):152-162. [doi: 10.1177/1460458212462151] [Medline: 23715214]
28. Wlodarczyk D, Chylińska J, Lazarewicz M, Rzadkiewicz M, Jaworski M, Adamus M, et al. Enhancing doctors' competencies in communication with and activation of older patients: the Promoting Active Aging (PRACTA) computer-based intervention study. J Med Internet Res 2017 Feb 22;19(2):e45 [FREE Full text] [doi: 10.2196/jmir.6948] [Medline: 28228370]
29. Knowles M. The Modern Practice of Adult Education. New York: New York Association Press; 1970.
30. Ashby L, Travis S. Teach yourself how to teach an older patient. RN 1988 Apr;51(4):25-28. [Medline: 3358067]
31. Simon SJ. The impact of culture and gender on web sites. SIGMIS Database 2001 Dec 31;32(1):18. [doi: 10.1145/506740.506744]
32. Penard T, Poussing N, Mukoko B, Tamokwe Piaptie GB. Internet adoption and usage patterns in Africa: evidence from Cameroon. Technol Soc 2015 Aug;42:71-80. [doi: 10.1016/j.techsoc.2015.03.004]
33. Wasserman IM, Richmond-Abbott M. Gender and the Internet: causes of variation in access, level, and scope of use. Social Science Q 2005 Mar;86(1):252-270. [doi: 10.1111/j.0038-4941.2005.00301.x]
34. Degeratu AM, Rangaswamy A, Wu J. Consumer choice behavior in online and traditional supermarkets: the effects of brand name, price, and other search attributes. Int J Res Mark 2000 Mar;17(1):55-78. [doi: 10.1016/S0167-8116(00)00005-7]
35. Nayak LU, Priest L, White AP. An application of the technology acceptance model to the level of Internet usage by older adults. Univ Access Inf Soc 2010 Feb 16;9(4):367-374. [doi: 10.1007/s10209-009-0178-8]
36. Wood E, Lanuza C, Baciu I, MacKenzie M, Nosko A. Instructional styles, attitudes and experiences of seniors in computer workshops. Educ Gerontol 2010 Sep 07;36(10-11):834-857. [doi: 10.1080/03601271003723552]
37. Ramón-Jerónimo MA, Peral-Peral B, Arenas-Gaitán J. Elderly persons and Internet use. Soc Sci Comput Rev 2013 Feb 07;31(4):389-403. [doi: 10.1177/0894439312473421]
38. Hashmi N, Manzoor I, Naser F. Computer use in medical students. Computer 2012;78(18):22.
39. Buse CE. When you retire, does everything become leisure? Information and communication technology use and the work/leisure boundary in retirement. New Media Soc 2009 Oct 06;11(7):1143-1161. [doi: 10.1177/1461444809342052]
40. Chiu C, Hu Y, Lin D, Chang F, Chang C, Lai C. The attitudes, impact, and learning needs of older adults using apps on touchscreen mobile devices: results from a pilot study. Comput Hum Behav 2016 Oct;63:189-197. [doi: 10.1016/j.chb.2016.05.020]
41. Lee C, Myrick R, D'Ambrosio L, Coughlin J, de Weck OL. Older adults' experiences with technology: learning from their voices. In: Communications in Computer and Information Science. Berlin: Springer; 2013:251-255.
42. Lee C, Coughlin JF. PERSPECTIVE: Older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J Prod Innov Manag 2014 Jun 03;32(5):747-759. [doi: 10.1111/jpim.12176]
43. Huber L, Watson C. Technology: education and training needs of older adults. Educ Gerontol 2013 Sep 06;40(1):16-25. [doi: 10.1080/03601277.2013.768064]
44. Charness N, Boot W. Aging and information technology use. Curr Dir Psychol Sci 2009 Oct;18(5):253-258. [doi: 10.1111/j.1467-8721.2009.01647.x]
45. Campbell R. Older women and the internet. J Women Aging 2004;16(1-2):161-174. [doi: 10.1300/J074v16n01_11] [Medline: 15149930]
46. Becker SA. A study of web usability for older adults seeking online health resources. ACM Trans Comput-Hum Interact 2004 Dec 01;11(4):387-406. [doi: 10.1145/1035575.1035578]
47. Baltes BP, Baltes MM. Psychological perspectives on successful aging: the model of selective optimization with compensation. In: Baltes PB, Baltes MM, editors. Successful Aging: Perspectives From the Behavioral Sciences. Cambridge: Cambridge University Press; 1990:1-34.
48. Hanson V. Technology skill and age: what will be the same 20 years from now? Univ Access Inf Soc 2011 Apr 20;10(4):443-452. [doi: 10.1007/s10209-011-0224-1]
49. Kim J, Lee H, Christensen M, Merighi J. Technology access and use, and their associations with social engagement among older adults: do women and men differ? J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):836-845. [doi: 10.1093/geronb/gbw123] [Medline: 28073816]

Edited by G Eysenbach; submitted 24.01.17; peer-reviewed by D Tao, YH Chang; comments to author 05.06.17; revised version received 13.09.17; accepted 30.09.17; published 03.11.17.

Please cite as:
Chiu CJ, Liu CW
Understanding Older Adult's Technology Adoption and Withdrawal for Elderly Care and Education: Mixed Method Analysis from National Survey
J Med Internet Res 2017;19(11):e374
URL: http://www.jmir.org/2017/11/e374/
doi:10.2196/jmir.7401
PMID:29101093

Original Paper

Bioimpedance Alerts from Cardiovascular Implantable Electronic Devices: Observational Study of Diagnostic Relevance and Clinical Outcomes

Christophe JP Smeets[1,2,3], MSc; Julie Vranken[1,3], MSc; Jo Van der Auwera[1], MSc; Frederik H Verbrugge[2], MD, PhD; Wilfried Mullens[2], MD, PhD; Matthias Dupont[2], MD, PhD; Lars Grieten[1], MSc, PhD; Hélène De Cannière[1,2,3], MSc; Dorien Lanssens[1,3], MSc; Thijs Vandenberk[1,3], MSc; Valerie Storms[1], MSc, PhD; Inge M Thijs[1,3], MSc, PhD; Pieter M Vandervoort[1,2,3], MD

[1]Mobile Health Unit, Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium

[2]Department of Cardiology, Ziekenhuis Oost-Limburg, Genk, Belgium

[3]Future Health Department, Ziekenhuis Oost-Limburg, Genk, Belgium

Corresponding Author:
Christophe JP Smeets, MSc
Department of Cardiology
Ziekenhuis Oost-Limburg
Schiepse Bos 6
Genk,
Belgium
Phone: 32 089 32 15 26
Email: christophe.smeets@uhasselt.be

Abstract

Background: The use of implantable cardioverter-defibrillators (ICDs) and cardiac resynchronization therapy (CRT) devices is expanding in the treatment of heart failure. Most of the current devices are equipped with remote monitoring functions, including bioimpedance for fluid status monitoring. The question remains whether bioimpedance measurements positively impact clinical outcome.

Objective: The aim of this study was to provide a comprehensive overview of the clinical interventions taken based on remote bioimpedance monitoring alerts and their impact on clinical outcome.

Methods: This is a single-center observational study of consecutive ICD and CRT patients (n=282) participating in protocol-driven remote follow-up. Bioimpedance alerts were analyzed with subsequently triggered interventions.

Results: A total of 55.0% (155/282) of patients had an ICD or CRT device equipped with a remote bioimpedance algorithm. During 34 (SD 12) months of follow-up, 1751 remote monitoring alarm notifications were received (2.2 per patient-year of follow-up), comprising 2096 unique alerts (2.6 per patient-year of follow-up). Since 591 (28.2%) of all incoming alerts were bioimpedance-related, patients with an ICD or CRT including a bioimpedance algorithm had significantly more alerts (3.4 versus 1.8 alerts per patient-year of follow-up, $P<.001$). Bioimpedance-only alerts resulted in a phone contact in 91.0% (498/547) of cases, which triggered an actual intervention in 15.9% (87/547) of cases, since in 75.1% (411/547) of cases reenforcing heart failure education sufficed. Overall survival was lower in patients with a cardiovascular implantable electronic device with a bioimpedance algorithm; however, this difference was driven by differences in baseline characteristics (adjusted hazard ratio of 2.118, 95% CI 0.845-5.791). No significant differences between both groups were observed in terms of the number of follow-up visits in the outpatient heart failure clinic, the number of hospital admissions with a primary diagnosis of heart failure, or mean length of hospital stay.

Conclusions: Bioimpedance-only alerts constituted a substantial amount of incoming alerts when turned on during remote follow-up and triggered an additional intervention in only 16% of cases since in 75% of cases, providing general heart failure education sufficed. The high frequency of heart failure education that was provided could have contributed to fewer heart failure–related hospitalizations despite significant differences in baseline characteristics.

(J Med Internet Res 2017;19(11):e393) doi:10.2196/jmir.8066

KEYWORDS

defibrillators, implantable; cardiac resynchronization therapy; telemedicine; electric impedance; algorithms; call centers

Introduction

Cardiac resynchronization therapy (CRT) devices and implantable cardioverter-defibrillators (ICDs) are guideline-recommended treatments for heart failure with reduced ejection fraction, left bundle branch block, very wide QRS complex (>150 ms), or sudden cardiac death [1,2]. The use of these cardiovascular implantable electronic devices (CIEDs) is rapidly increasing with 51,274 and 85,289 patients, respectively, receiving a CRT or ICD device in Europe in 2013 [3]. Remote follow-up of this group is slowly finding its way into routine clinical practice since it may hold major advantages for patients, health care workers, and society [4]. Over the last decade, several CIED manufacturers have marketed thoracic impedance measurement algorithms integrated into their devices. Changes in bioimpedance measurements reflect changes in intrathoracic fluid status and are evaluated based on a vendor-specific computer algorithm. Early investigations reported an inverse correlation with pulmonary capillary wedge pressure and fluid balance [5] and a higher sensitivity and lower unexplained detection rate compared to acute weight changes [6]. In addition, a decrease in bioimpedance happened even before clinical manifestation of heart failure worsening and before hospital admission for fluid overload [5]. These algorithms are therefore very promising for the early detection of impending decompensated heart failure and enable the possibility to adjust treatment strategies in order to prevent heart failure hospitalization [7,8]. Despite early investigations showing promising results, larger randomized trials have revealed disappointing outcomes [9-12]. A shortcoming in current studies is the lack of standardization and information as to what clinical actions are coupled to remote bioimpedance alerts, making it difficult to draw conclusions on its clinical impact.

In 2010, a dedicated remote follow-up program of heart failure patients with a CIED was started at Ziekenhuis Oost-Limburg (Genk, Belgium). Dedicated nurses, trained in electrophysiology, device follow-up, and heart failure pathophysiology, review all incoming alerts in a systematic and standardized manner with automatic interventions triggered by protocol-based, guideline-recommended care [13,14]. In this observational registry study, we closely analyzed all bioimpedance alerts and subsequent triggered interventions from OptiVol and OptiVol 2.0 (Medtronic PLC) and CorVue (St. Jude Medical LLC) CIEDs using the default alert settings, and we studied their impact on clinical outcome. As such, the current research builds on previous studies since these lack this level of detail.

Methods

Study Design

This is an observational registry study of ICD and CRT patients from a single tertiary care center (Ziekenhuis Oost-Limburg, Genk, Belgium) implanted with the devices between February 2010 and May 2013. Since February 2010, all patients receiving either an ICD or CRT device with remote monitoring capabilities were asked to participate voluntarily in a remote follow-up program. The type of CIED that was implanted was solely based on the device's therapeutic capabilities (right ventricular pacing, biventricular pacing, and antitachycardia treatment) and was left to the discretion of the treating physician. For this analysis, only patients enrolled in remote follow-up within 6 months after device implantation are included. Patient baseline information is collected at the time of device implantation. All participants provided written informed consent and were followed until February 1, 2015. The study complies with the Declaration of Helsinki, and the study protocol was approved by the local committee on human research.

Remote Follow-Up and Alerts

A vendor-specific transmission device, usually installed in the patient's bedroom, collected disease- and device-related data from the CIED that was transmitted to an online database accessible to the multidisciplinary heart failure team. All alerts were interpreted on weekdays by dedicated nurses trained in electrophysiology, device follow-up, and heart failure pathophysiology; notifications received during weekends were read on Monday. Daily alert transmissions were generated when predefined alarm thresholds were crossed. Besides alert transmissions, each device was programmed to send a scheduled transmission report monthly. Alerts were categorized according to their nature into technical (missed scheduled transmission and technical device problems) or clinical (rhythm, bioimpedance, and miscellaneous [changes in daily activity, heart failure management, etc]) alerts. Our study focuses on all bioimpedance-related alerts.

Bioimpedance Measurements

Since bioimpedance is measured from the electrode lead to the device can, any thoracic fluid change including vascular, interstitial, or alveolar fluid results in a change in its value. Therefore, bioimpedance measurements are not specific to one disease.

In 2004, Medtronic was the first company to introduce a bioimpedance algorithm, known as the OptiVol algorithm, in its CIEDs. For the OptiVol algorithm, bioimpedance is measured in a semicontinuous way every 20 minutes from 12 AM to 5 PM. The algorithm starts 34 days postimplant and generates 2 separate graphs: one displays the raw bioimpedance data and the other indicates the accumulated change between the daily bioimpedance measurements and a dynamic reference impedance. The latter one, called the OptiVol fluid index, triggers an alarm when a predefined threshold is met, by default set at 60Ω [5]. The OptiVol fluid index graph may indicate an event, while the raw bioimpedance graph may indicate the severity of the event [15,16]. In 2010, Medtronic launched OptiVol 2.0, an updated version of the initial bioimpedance algorithm. The updated version is intended to lower the number of false positive alerts. Alterations include a faster changing reference after initialization, a slower accumulating fluid index for an initial duration of the event in patients with higher day-to-day variability in impedance, and a fluid index which

accumulates only over the last 30 days [15]. An initial study has reported a 40% decrease in unexplained detections [17]. In our study, the default OptiVol threshold settings were used.

In 2009, St Jude Medical introduced its own bioimpedance algorithm, known as CorVue. There are some fundamental differences compared to the OptiVol algorithms. CorVue also measures intrathoracic impedances in a semicontinuous way every 2 hours around the clock. In addition, depending on the type of lead, the impedance is measured in one (ie, unipolar leads) or multiple vectors. Within the first 2 weeks, the algorithm starts to build a reference impedance which is a long-term moving average (ie, over the last 144 or 168 measurements for CRT or ICD devices, respectively). Afterward, a short-term moving average (ie, over the last 12 measurements) of multivector impedance measurements builds the daily impedance. A bioimpedance alert is triggered when the daily impedance is lower than the reference impedance for a programmable duration known as the congestion trigger (ie, nominal 13 days for ICDs, 14 days for CRT-D, and 16 days for CRT-P) [15]. In our study, the default congestion trigger settings were used.

The presence of a bioimpedance algorithm is dependent on the CIED manufacturer: most Medtronic and St Jude devices are equipped with a bioimpedance algorithm and generate bioimpedance alerts, while Biotronik and Boston Scientific devices do not generate any bioimpedance alerts. In our study, the choice of CIED brand implanted in a particular patient is completely random. Therefore, the presence or absence of a bioimpedance algorithm in this patient population is also randomly assigned.

Intervention Protocol for Bioimpedance Alerts

All incoming bioimpedance alerts were deemed to be of potential clinical relevance and resulted in a phone contact between the interpreting nurse and the patient. In exceptional cases, where the patient had an in-hospital check-up very recently or had one planned in the near future, a phone contact was not initiated. A custom-made heart failure questionnaire was used to identify potential causes for the bioimpedance alert [14,18]. Additional questions could be asked at the discretion of the health care worker in order to gain better insight. Appropriate feedback and general heart failure education (ie, stress the importance of the conservation of a salt-free diet and fluid restrictions) were always provided. Further action was protocol-driven in consultation with a dedicated heart failure specialist.

Outpatient Follow-Up

Patients enrolled in remote follow-up visit the outpatient cardiology clinic for device and clinical heart failure follow-up at 6 weeks after implantation and subsequently every 6 months with a minimum of 2 visits per year as per standard practice in our institution. Patients in this study were followed until death, exclusion from remote follow-up, heart transplantation, or February 1, 2015, whichever came first.

Statistical Analysis

Demographic and functional characteristics were compared using descriptive statistics. Continuous variables are expressed as mean and standard deviation (SD) if normally distributed or median and interquartile range (IQR) otherwise. Survival curves were constructed according to the Kaplan Meier method, with the log-rank test used for comparison among groups. Unadjusted and adjusted hazard ratios (HRs) were calculated by Cox regression analysis with Firth's penalized likelihood correction. To define statistical differences between both groups, the independent samples Student *t* test and Mann-Whitney U test were used for normally and not normally distributed continuous variables, respectively, and the chi-square test and Fisher exact test were used accordingly for categorical variables. To define statistical differences between the different bioimpedance algorithms, the Kruskal Wallis test was used. The significance level for tests was 2-sided with alpha=.05. All statistical analyses were performed using IBM SPSS Statistics 24.0 (IBM Corp); SAS 9.4 (SAS Institute Inc) was used for Cox regression with Firth's penalization.

Results

Study Population

From a total of 506 patients with a CIED implanted during the study period, 110 patients were excluded due to the presence of a cardiac resynchronization therapy pacemaker (CRT-P) device without remote monitoring capabilities, 82 patients were excluded because the remote monitoring program was started more than 6 months after device implantation, 22 patients refused study participation, and 10 patients were excluded due to follow-up in another center. The final study population consisted of 282 patients: 155 (55.0%) patients with a CIED equipped with a bioimpedance algorithm ($CIED^{+}$) and 127 (45.0%) patients with a CIED without an available bioimpedance algorithm ($CIED^{-}$) (Figure 1). Of 282 devices, 110 (39.0%) Medtronic, 105 (37.2%) St. Jude Medical, 61 (21.6%) Biotronik, and 6 (2.1%) Boston Scientific CIEDs were implanted.

Respectively, 26.4% (41/155), 43.2% (67/155), and 30.3% (47/155) of patients in the $CIED^{+}$ population had a device implanted with Optivol, Optivol 2.0, and CoreVue algorithm. The median time interval between CIED implantation and start of remote follow-up was 1 day (IQR 1 to 2 days), with 81.9% (231/282) of patients included within 1 week. Patients were followed for 34 (SD 12) months leading to 801 cumulative patient-years of follow-up. The number of follow-up visits in the outpatient device clinic was 3.25 per patient-year of follow-up, of which 93% were elective and 7% were triggered by remote monitoring. Baseline characteristics of the study population at the time of implantation are shown in Table 1.

Remote Follow-Up Notifications, Alerts, and Interventions

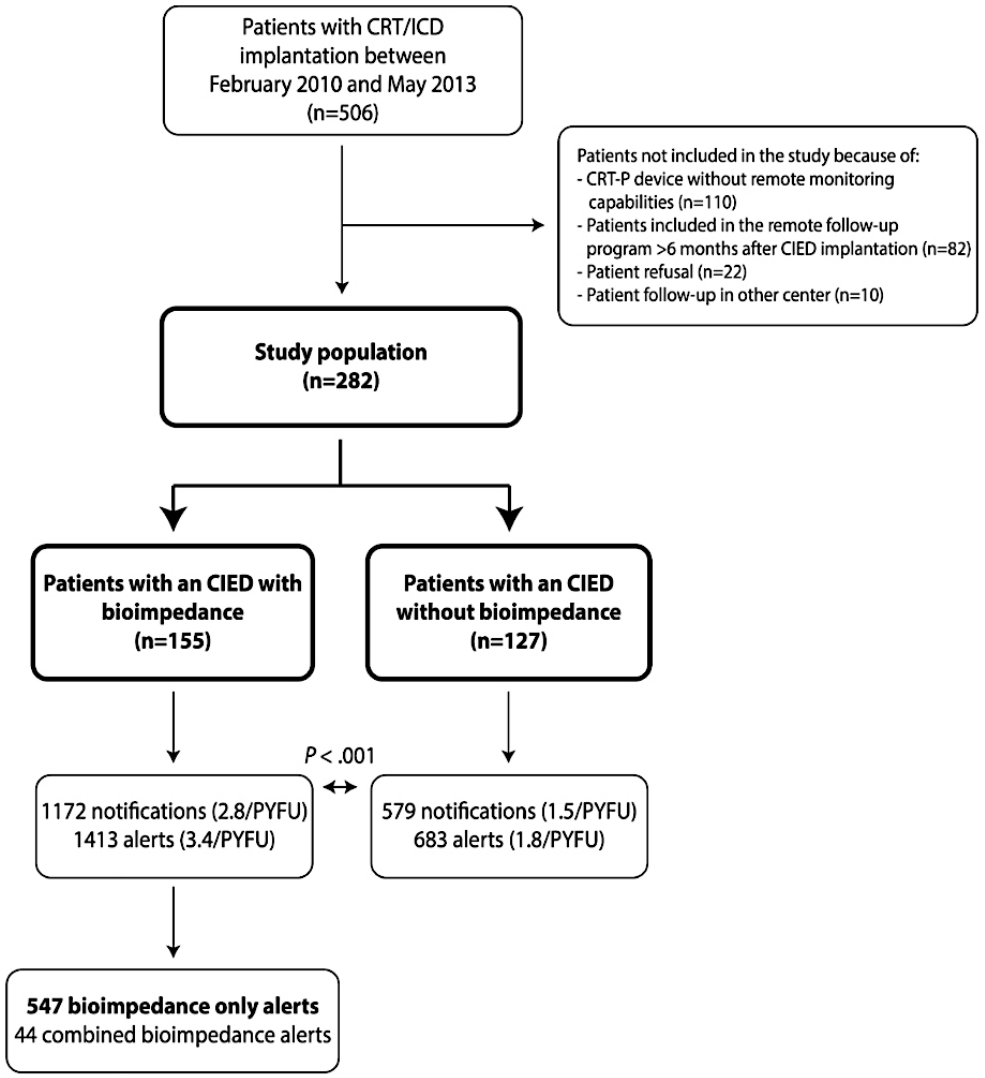
During follow-up, the clinical call center handled 1751 remote monitoring notifications. Since a notification can contain multiple alerts, a total of 2096 unique alerts were received (ie, 2.6 alerts per patient-year of follow-up). Patients with a $CIED^{+}$

had significantly more alerts than those with a CIED^{-}: 1413 (67.41%, 3.4 per patient-year of follow-up) and 683 (32.59%, 1.8 per patient-year of follow-up), respectively, $P<.001$. The amount of technical and arrhythmia alerts was similar in both patient groups. The higher number of alerts in the CIED^{+} population can be entirely attributed to bioimpedance alerts. The distribution of the different alert categories among both groups is shown in Figure 2.

Bioimpedance Notifications and Interventions

During follow-up, 591 notifications including a bioimpedance threshold crossing were received for 111 of 155 (71.6%) patients. In 44 of these notifications, 1 or more additional alerts were combined resulting in 547 bioimpedance-only notifications. In 498 (91.0%) of bioimpedance-only notifications, the patient was contacted by phone and a standardized heart failure questionnaire was used. In 9.0% (49/547) of cases, a phone contact was not initiated since the patient had an in-hospital check-up very recently or had one planned in the near future. In 75.1% (411/547) of bioimpedance-only notifications, only general heart failure education was given. An additional intervention was triggered in 15.9% (97/547) of cases (Figure 3, left). In total, 97 interventions were performed (Figure 3, right), including medication changes in 50% (48/97) of cases, referral to the general practitioner or cardiologist in 27% (26/97) and 23% (22/97) of cases, respectively, and in 1% (1/97) the patient was asked to visit the emergency room. A combination of different interventions for 1 bioimpedance alert is also possible.

Figure 1. Flowchart of the study.



CRT, cardiac resynchronization therapy; ICD, implantable cardioverter defibrillator; CIED, cardiovascular implantable electronic devices; PYFU, patient-years of follow-up.

Table 1. Baseline characteristics of the study population (n=282).

Variables	CIED[a] with bioimpedance n=155	CIED without bioimpedance n=127	*P* value
Age, years, mean (SD)	72 (12)	70 (13)	.12
BMI[b], mean (SD)	27 (5)	28 (6)	.54
Male gender, n (%)	123 (79.4)	108 (85.0)	.22
ICD[c], n (%)	31 (20.0)	59 (46.5)	<.001
CRT-D[d], n (%)	102 (65.8)	57 (44.9)	<.001
CRT-P[e], n (%)	22 (14.2)	11 (8.7)	.15
Bioimpedance algorithm, n (%)			
OptiVol	41 (26.4)	—	
OptiVol 2.0	67 (43.2)	—	
CorVue	47 (30.3)	—	
NYHA[f,g] functional class, n (%)			.42
Class II	17 (15.6)	14 (23.7)	
Class III	90 (82.6)	43 (72.9)	
Left ventricular ejection fraction, %, mean (SD)	31 (12)	34 (12)	.01
QRS width, ms, mean (SD)	145 (31)	127 (32)	<.001
Heart failure etiology, n (%)			
Ischemic heart disease	86 (55.5)	86 (67.7)	.04
Dilated	13 (8.3)	5 (3.9)	.13
Valvular	3 (1.9)	1 (0.8)	.63
Hypertrophic	5 (3.2)	3 (2.4)	.73
Toxic	1 (0.6)	2 (1.6)	.59
Idiopathic	40 (25.8)	17 (13.4)	.01
Other etiology or no heart failure	7 (4.5)	13 (10.2)	.06
Comorbidities and risk factors, n (%)			
Valvular surgery	18 (11.6)	11 (8.7)	.42
Atrial fibrillation	62 (40.0)	51 (40.2)	.98
Chronic obstructive pulmonary disease	23 (14.8)	16 (12.6)	.59
Chronic kidney disease	42 (27.1)	31 (24.4)	.61
Cerebrovascular accident	13 (8.4)	10 (7.9)	.88
Diabetes	32 (20.6)	25 (19.7)	.84
Family history of cardiovascular disease	34 (21.9)	40 (31.5)	.07
Arterial hypertension	59 (38.1)	57 (44.9)	.25
Hypercholesterolemia	59 (38.1)	48 (37.8)	.96
Smoking	151 (97.4)	122 (96.1)	.74
Medication use, n (%)			
Renin-angiotensin system blocker	125 (80.6)	102 (80.3)	.94
Beta-blocker	144 (92.9)	109 (85.8)	.05
Spironolactone	105 (67.7)	68 (53.5)	.02
Loop diuretic	76 (49.0)	48 (37.8)	.06
Digoxin	23 (14.8)	18 (14.2)	.88

Variables	CIED[a] with bioimpedance n=155	CIED without bioimpedance n=127	*P* value
Statin	85 (54.8)	82 (64.6)	.10
Calcium channel blockers	8 (5.2)	15 (11.8)	.04
Antidiabetic medication	27 (17.4)	22 (17.3)	.98

[a]CIED: cardiovascular implantable electronic device.

[b]BMI: body mass index.

[c]ICD: implantable cardioverter-defibrillator.

[d]CRT-D: cardiac resynchronization therapy defibrillator.

[e]CRT-P: cardiac resynchronization therapy pacemaker.

[f]NYHA: New York Heart Association.

[g]CIED with bioimpedance group (n=109), CIED without bioimpedance group (n=59).

Figure 2. Frequency of alert categories with the number of alerts per patient-year of follow-up for patients with a cardiovascular implantable electronic device with or without a bioimpedance algorithm. Disease-related alerts are marked in blue color tints and technical-related alerts in green color tints.

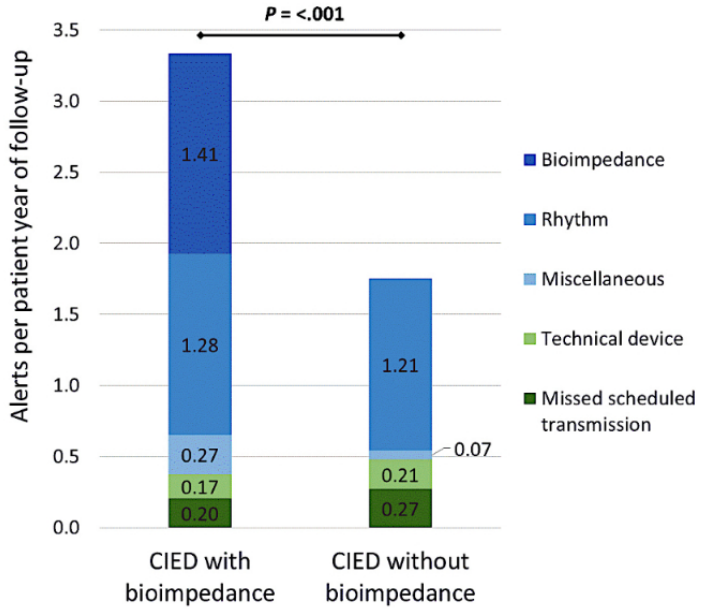
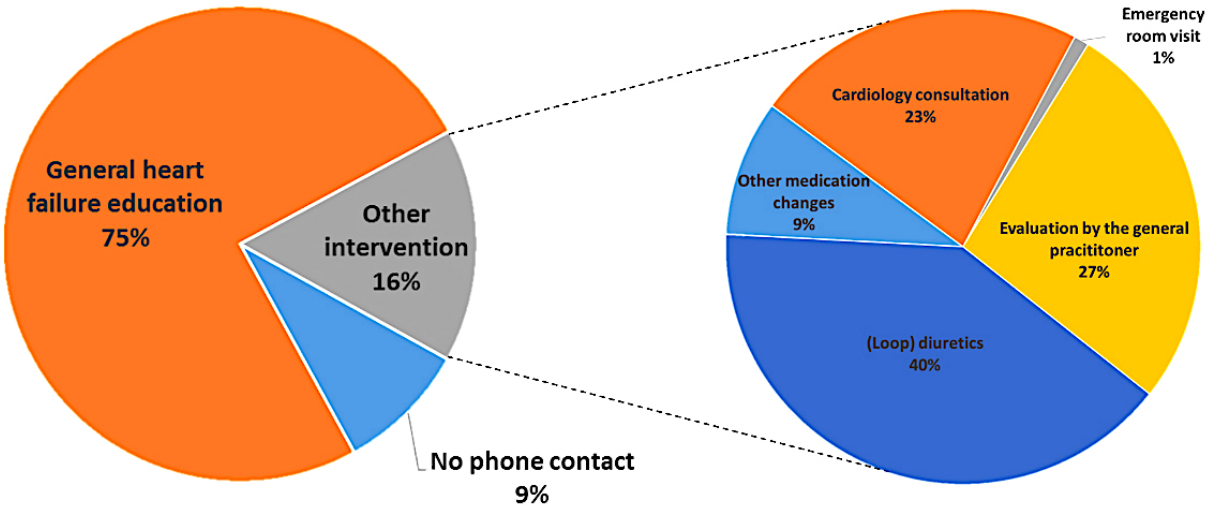


Figure 3. Overview of the interventions triggered during remote follow-up in the case of a bioimpedance-only alert.



Different Bioimpedance Algorithms

There was a statistically significant difference in number of bioimpedance alerts per patient-year between the different bioimpedance algorithms (χ^2=12.643, *P*=.002) (Figure 4). The updated OptiVol 2.0 algorithm (0.79 alerts per patient-year of follow-up) triggered significantly fewer bioimpedance alerts than OptiVol (1.67 alerts per patient-year of follow-up; *P*=.02) and CorVue (1.97 alerts per patient-year of follow-up; *P*=.005). No differences were observed concerning the distribution of interventions triggered by the different bioimpedance algorithms.

Clinical Outcome

At mean time of follow-up (ie, 34 months), 26 patients had died, leading to an overall survival rate of 90.8%. Seven patients died in the CIED$^-$ population compared to 19 in the CIED$^+$ population, leading to all-cause survival rates of 94.5% and 87.7%, respectively (*P*=.047) (Figure 5A). Most deaths were due to cardiovascular causes, with 6 in the CIED$^-$ population compared to 15 in the CIED$^+$ population (*P*=.10). No significant differences in survival rate were observed for the different bioimpedance algorithms.

At mean time of follow-up, 40 patients were hospitalized with a primary diagnosis of heart failure and hence 85.8% (242/282) of patients were free from heart failure–related hospitalization. No significant difference was observed between both groups (23/155, 85.2%, for CIED$^+$ versus 17/127, 86.6%, for CIED$^-$ at mean time of follow-up, *P*=.76) (Figure 5B) or for the different bioimpedance algorithms (*P*=.95) (Figure 5C).

No significant differences were observed between both groups with respect to the number of elective follow-up visits in the outpatient heart failure clinic (*P*=.45) or the number of cardiac-related hospital admissions (*P*=.32). For those who had at least 1 cardiac-related hospital admission, median length of hospital stay was 6 (IQR 3 to 14) days. There was no significant difference for length of hospital stay between both groups or for the different bioimpedance algorithms.

Figure 4. Overview of the amount of remote monitoring bioimpedance alerts per patient-year of follow-up triggered by the different bioimpedance algorithms.

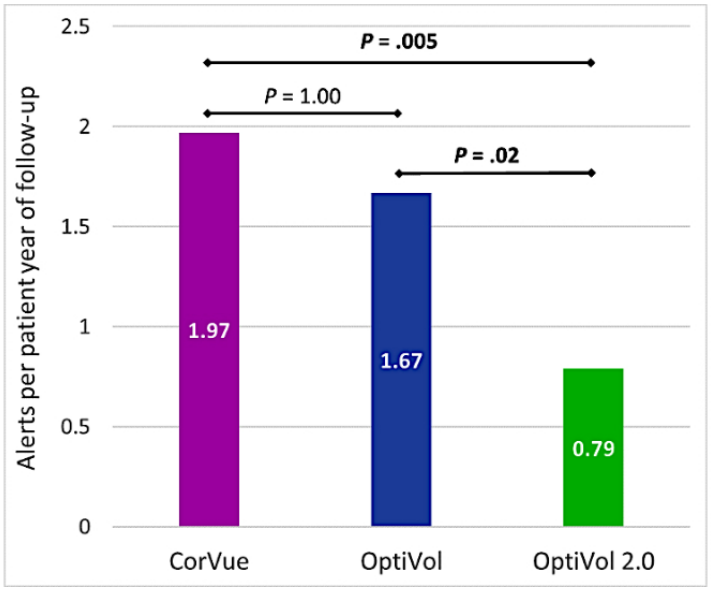


Figure 5. (A) Probability of survival for patients with a cardiovascular implantable electronic device (CIED) with or without a bioimpedance algorithm, (B) Freedom from hospital admission with a primary diagnosis of heart failure for patients with a CIED with or without a bioimpedance algorithm, (C) Freedom from hospital admission with a primary diagnosis of heart failure for the different bioimpedance algorithms.

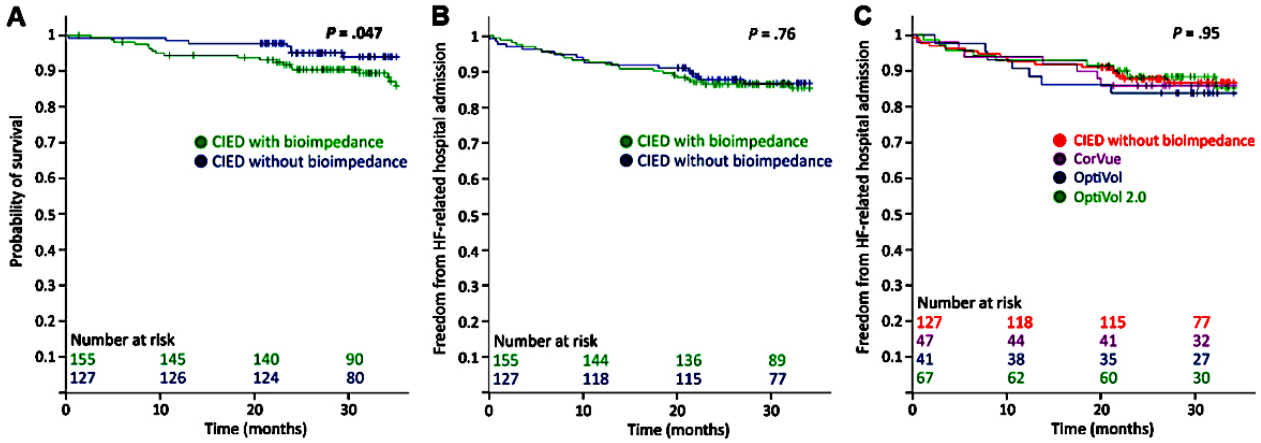


Table 2. Cox regression analysis with Firth's penalization for clinical outcome measures.

Variables	Unadjusted hazard ratio			Adjusted hazard ratio[a]		
	HR[b]	95% CI	*P* value	HR	95% CI	*P* value
All-cause mortality	2.342	1.029-5.996	.047	2.118	0.845-5.791	.13
Cardiovascular survival	2.168	0.881-6.082	.10	2.335	0.852-7.020	.12
Heart failure hospitalization	1.103	0.592-2.097	.76	1.284	0.655-2.562	.47

[a]Hazard ratios were adjusted for significant differences in baseline characteristics including implantable cardioverter-defibrillator use, cardiac resynchronization therapy defibrillator use, left ventricular ejection fraction, QRS width, ischemic etiology of heart failure and spironolactone use, and clinically relevant parameters including age, gender, and loop diuretic use.

[b]HR: hazard ratio.

Table 2 provides an overview of the unadjusted and adjusted Cox regression analysis with Firth's penalization. Presence of bioimpedance algorithms in the CIED resulted in a nonsignificant adjusted hazard ratio of 2.118 (95% CI 0.845-5.791) for all-cause death and 2.335 (95% CI 0.852-7.020) for cardiovascular death. Multivariate analysis indicated that age, ejection fraction, and QRS time contribute to the observed difference in survival rate between both groups.

Discussion

Principal Findings

Although many newly implanted CIEDs have a built-in bioimpedance algorithm, it remains unclear whether bioimpedance measurements contribute to improved clinical outcome when incorporated in a standardized heart failure care path including remote follow-up. In this paper, we present a comprehensive overview of bioimpedance alerts and subsequent triggered interventions in patients with either a CRT or ICD device enrolled in a dedicated, protocol-driven, remote follow-up program in a single Belgian tertiary care center.

Major insights include the following:

1. Patients with a CIED equipped with a bioimpedance algorithm have significantly more remote monitoring notifications
2. In 75% of bioimpedance-only alerts, reenforcing heart failure education was the only action taken; in 16% of cases, an additional intervention was triggered; and in 9%, the patient was not contacted
3. For the different bioimpedance algorithms, significant differences were observed for the number of bioimpedance alerts but not for triggered interventions or clinical outcome

Although patients with a CIED equipped with a bioimpedance algorithm have a significantly lower survival rate, driven by differences in baseline characteristics, there was no difference in heart failure–related hospitalizations.

An important observation that merits further attention is the high number of patients with bioimpedance-only alerts who were contacted and given heart failure education only without any additional interventions. This may indicate a high sensitivity but low specificity of these alerts to detect emerging congestion as well as the existence of a temporal lag between a bioimpedance alert and clinical manifestation of heart failure worsening [5,7-12,19-21]. Therefore, a possible explanation for the rather low number of additional interventions could be that due to these bioimpedance alerts, patients are contacted in the early phase of emerging congestion. By the fact that patients are contacted and general heart failure education is repeated, it is possible that their perception of disease awareness strengthens (ie, importance of fluid and salt restriction, heart failure medication intake, and physical activity), avoiding further worsening of congestion.

The updated OptiVol 2.0 algorithm triggered significantly fewer bioimpedance alerts than the other two bioimpedance algorithms. This corresponds to literature, where a 40% decrease in unexplained bioimpedance alerts was observed [17]. However, no changes in intervention strategy or impact on clinical outcome compared to the other bioimpedance algorithms was observed. Although improvements to bioimpedance algorithms have already been made, this could still indicate that intrathoracic impedance is currently wrongly measured, handled, or interpreted. In the majority of cases, the bioimpedance alarm threshold is set to default. These thresholds should be individually adjusted in order to improve sensitivity and specificity rates, as suggested by previous research [7,8,15]. Another possibility to improve intrathoracic impedance measurements is controlling the circumstances in which measurements are performed. Currently, measurements are performed under different circumstances throughout the day. Since bioimpedance measurements are also influenced by motion and body posture, reliability could be improved by a lower number of measurements that are all performed under the same posture (eg, during the night when lying in a particular posture). Finally, instead of alerts triggered by bioimpedance crossings alone, integration with other parameters currently monitored by implantable electrical devices (eg, patient activity, heart rate variability, average ventricular rate) will provide a more efficient tool to predict heart failure worsening [18].

Patients with a CIED$^+$ showed a lower overall survival rate. However, this difference can be explained by differences in baseline characteristics between both groups rather than the presence of a bioimpedance algorithm. Indeed, multivariate analysis indicated a significant hazard ratio for age, ejection fraction, and QRS time and a nonsignificant adjusted hazard ratio for the presence of a bioimpedance algorithm in the CIED. Concerning heart failure–related hospitalizations, both groups showed similar results. This could indicate a potential benefit of bioimpedance algorithms on clinical outcome since, based on baseline characteristics, one would expect a higher number

of heart failure–related hospitalizations in the CIED$^+$ group. This could mean that the high frequency of heart failure education provided in cases of a bioimpedance alert could have prevented heart failure–related hospitalizations.

When reviewing available literature, it is clear that the success rate of remote monitoring is strongly dependent on optimal workflow with standardized protocols and appropriate feedback loops. The DOT-HF (Diagnostic Outcome Trial in Heart Failure) trial [9], where patients received an audible alert in case of a bioimpedance crossing, showed that providing wrong feedback can even induce an increase in hospital admissions. In the LIMIT-CHF (Lung Impedance Monitoring in Treatment of Chronic Heart Failure) trial [15], bioimpedance alerts triggered empirical changes in diuretic dose, which did not significantly prevent heart failure–related hospitalizations. Moreover, the OPTILINK-HF (Optimization of Heart Failure Management using OptiVol Fluid Status Monitoring and CareLink) trial [12] employed a similar approach of protocol-driven remote monitoring as used in our center and reported no significant improvements in clinical outcome. Shortcomings in this trial were the single-parameter follow-up, suboptimal data transmission, and the absence of a centralized monitoring team. In our study, a multiparameter approach was used and all incoming alerts were handled in a standardized way. In addition, remote follow-up in our center is performed by a small team of dedicated heart failure nurses who have close personal contact with the patients. Furthermore, our nurses are operating from inside our tertiary care center and hence have daily contact with the treating physician. This approach facilitates protocol standardization and has already been shown to be effective in the IN-TIME (Influence of Home Monitoring on Mortality and Morbidity in Heart Failure Patients With Impaired Left Ventricular Function) trial [22]. It is clear that there is not just one remote monitoring approach, but a high variability exists and hence each approach needs to be assessed on its individual merit. In our study, the question remains to what extent reinforcement of heart failure education is crucial in remote bioimpedance monitoring and impacts clinical outcome.

Study Limitations

This study should be interpreted in the light of some limitations. First, this is a relatively small single center observational registry study with classic limitations associated with this type of study design, thereby making the study results mainly hypothesis-generating. Although a nonsignificant adjusted hazard ratio was obtained for the presence of bioimpedance algorithms in the CIED, a potential power problem can be present in the all-cause survival analysis due to the rather low sample size and low event rates. Next, since the presence of a bioimpedance algorithm is dependent on CIED manufacturer and is therefore random, our study is a nonrandomized clinical trial. A possible selection bias could be present due to the device indication. However, the assignment of the type of CIED that was implanted was solely based on the device's therapeutic capabilities (right ventricular pacing, biventricular pacing, and antitachycardia treatment) and was left to the discretion of the treating physician. Other diagnostic information, for example, the presence or absence of bioimpedance algorithms in the CIED, was not taken into consideration when assigning the CIED type or brand. In the CIED$^+$ population, more CRT-D devices are present since these devices were first equipped with bioimpedance algorithms. In general, patients who receive a CRT-D device have more advanced heart failure than patients who receive an ICD device. This can explain the differences in baseline characteristics (eg, older population with a lower left ventricular ejection fraction and broader QRS complex). Finally, study inclusion was based on voluntary participation to remote follow-up. Therefore, one cannot exclude the possibility that enrolled patients were more motivated for follow-up with better expected compliance to therapies. However, the majority (>95%) of patients agree to remote follow-up, reducing the risk for selection bias.

Conclusion

In patients with a CIED with a bioimpedance algorithm, bioimpedance alerts constitute almost half (42%) of incoming alerts when turned on during remote follow-up. Repeating general heart failure education by phone sufficed in 75% of cases, and an additional intervention was performed in 16% of cases. The high frequency of heart failure education that was provided could have contributed to fewer heart failure–related hospitalizations despite significant differences in baseline characteristics. Future trials are needed to verify whether bioimpedance algorithms can only be used to trigger heart failure education or if they have an intrinsic value to change treatment strategies. In addition, future improvements in the way bioimpedance is measured, handled, or interpreted could further increase its clinical relevance. Before bioimpedance measurements can be widely implemented in clinical practice, larger multicenter randomized controlled trials are required.

Acknowledgments

This report is part of the Limburg Clinical Research Program UHasselt-ZOL-Jessa, supported by the foundation Limburg Sterk Merk, Hasselt University, Ziekenhuis Oost-Limburg, and Jessa Hospital. No specific funding was provided for publication of this report. The authors thank the nursing staff of the cardiac remote monitoring program (Wendy Ceyssens, Linda Jacobs, and Jan Vercammen) for their excellent support in the data collection.

Conflicts of Interest

None declared.

References

1. McMurray JJV, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Eur J Heart Fail 2012 Aug;14(8):803-869. [doi: 10.1093/eurjhf/hfs105] [Medline: 22828712]
2. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013 Oct 15;62(16):e147-e239 [FREE Full text] [doi: 10.1016/j.jacc.2013.05.019] [Medline: 23747642]
3. Raatikainen MJP, Arnar DO, Zeppenfeld K, Merino JL, Levya F, Hindriks G, et al. Statistics on the use of cardiac electronic devices and electrophysiological procedures in the European Society of Cardiology countries: 2014 report from the European Heart Rhythm Association. Europace 2015 Jan;17 Suppl 1:i1-i75. [doi: 10.1093/europace/euu300] [Medline: 25616426]
4. Mullens W, Grimm RA, Verga T, Dresing T, Starling RC, Wilkoff BL, et al. Insights from a cardiac resynchronization optimization clinic as part of a heart failure disease management program. J Am Coll Cardiol 2009 Mar 03;53(9):765-773 [FREE Full text] [doi: 10.1016/j.jacc.2008.11.024] [Medline: 19245967]
5. Yu C, Wang L, Chau E, Chan RH, Kong S, Tang M, et al. Intrathoracic impedance monitoring in patients with heart failure: correlation with fluid status and feasibility of early warning preceding hospitalization. Circulation 2005 Aug 9;112(6):841-848 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.104.492207] [Medline: 16061743]
6. Abraham WT, Compton S, Haas G, Foreman B, Canby RC, Fishel R, et al. Intrathoracic impedance vs daily weight monitoring for predicting worsening heart failure events: results of the Fluid Accumulation Status Trial (FAST). Congest Heart Fail 2011;17(2):51-55 [FREE Full text] [doi: 10.1111/j.1751-7133.2011.00220.x] [Medline: 21449992]
7. Vollmann D, Nägele H, Schauerte P, Wiegand U, Butter C, Zanotto G, et al. Clinical utility of intrathoracic impedance monitoring to alert patients with an implanted device of deteriorating chronic heart failure. Eur Heart J 2007 Aug;28(15):1835-1840 [FREE Full text] [doi: 10.1093/eurheartj/ehl506] [Medline: 17309902]
8. Ypenburg C, Bax JJ, van der Wall EE, Schalij MJ, van Erven L. Intrathoracic impedance monitoring to predict decompensated heart failure. Am J Cardiol 2007 Feb 15;99(4):554-557. [doi: 10.1016/j.amjcard.2006.08.066] [Medline: 17293202]
9. van Veldhuisen DJ, Braunschweig F, Conraads V, Ford I, Cowie MR, Jondeau G, DOT-HF Investigators. Intrathoracic impedance monitoring, audible patient alerts, and outcome in patients with heart failure. Circulation 2011 Oct 18;124(16):1719-1726 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.043042] [Medline: 21931078]
10. Conraads VM, Tavazzi L, Santini M, Oliva F, Gerritse B, Yu C, et al. Sensitivity and positive predictive value of implantable intrathoracic impedance monitoring as a predictor of heart failure hospitalizations: the SENSE-HF trial. Eur Heart J 2011 Sep;32(18):2266-2273 [FREE Full text] [doi: 10.1093/eurheartj/ehr050] [Medline: 21362703]
11. Lüthje L, Vollmann D, Seegers J, Sohns C, Hasenfuß G, Zabel M. A randomized study of remote monitoring and fluid monitoring for the management of patients with implanted cardiac arrhythmia devices. Europace 2015 Aug;17(8):1276-1281. [doi: 10.1093/europace/euv039] [Medline: 25983310]
12. Böhm M, Drexler H, Oswald H, Rybak K, Bosch R, Butter C, OptiLink HF Study Investigators. Fluid status telemedicine alerts for heart failure: a randomized controlled trial. Eur Heart J 2016 Nov 01;37(41):3154-3163. [doi: 10.1093/eurheartj/ehw099] [Medline: 26984864]
13. Duchenne J, Verbrugge FH, Dupont M, Vercammen J, Jacobs L, Grieten L, et al. Implementation of transmural disease management in patients admitted with advanced heart failure. Acta Cardiol 2014 Apr;69(2):145-154. [doi: 10.2143/AC.69.2.3017295] [Medline: 24783465]
14. Smeets CJP, Verbrugge FH, Vranken J, Van der Auwera J, Mullens W, Dupont M, et al. Protocol-driven remote monitoring of cardiac resynchronization therapy as part of a heart failure disease management strategy. Acta Cardiol 2017 Aug 14:1-10. [doi: 10.1080/00015385.2017.1363022] [Medline: 28803515]
15. Domenichini G, Rahneva T, Diab IG, Dhillon OS, Campbell NG, Finlay MC, et al. The lung impedance monitoring in treatment of chronic heart failure (the LIMIT-CHF study). Europace 2016 Mar;18(3):428-435. [doi: 10.1093/europace/euv293] [Medline: 26683599]
16. Clinicians practical guide for using OptiVol and other trends for managing heart failure patients. URL: http://www.medtronic.com/content/dam/medtronic-com-m/mdt/crdm/documents/clinic_pract_gd_optivol.pdf [accessed 2017-06-20] [WebCite Cache ID 6rMX0Tlcy]
17. Sarkar S, Hettrick DA, Koehler J, Rogers T, Grinberg Y, Yu C, et al. Improved algorithm to detect fluid accumulation via intrathoracic impedance monitoring in heart failure patients with implantable devices. J Card Fail 2011 Jul;17(7):569-576. [doi: 10.1016/j.cardfail.2011.03.002] [Medline: 21703529]
18. Mullens W, Oliveira LPJ, Verga T, Wilkoff BL, Tang WHW. Insights from internet-based remote intrathoracic impedance monitoring as part of a heart failure disease management program. Congest Heart Fail 2010;16(4):159-163 [FREE Full text] [doi: 10.1111/j.1751-7133.2010.00149.x] [Medline: 20662868]
19. Heist EK, Herre JM, Binkley PF, Van Bakel AB, Porterfield JG, Porterfield LM, et al. Analysis of different device-based intrathoracic impedance vectors for detection of heart failure events (from the Detect Fluid Early from Intrathoracic Impedance Monitoring study). Am J Cardiol 2014 Oct 15;114(8):1249-1256. [doi: 10.1016/j.amjcard.2014.07.048] [Medline: 25150135]

20. Evangelista LS, Dracup K, Doering LV. Treatment-seeking delays in heart failure patients. J Heart Lung Transplant 2000 Oct;19(10):932-938. [Medline: 11044686]
21. Yang X, Hua W, Ding L, Wang J, Zheng L, Li C, et al. OptiVol fluid index predicts acute decompensation of heart failure with a high rate of unexplained events. J Geriatr Cardiol 2013 Sep;10(3):253-257 [FREE Full text] [doi: 10.3969/j.issn.1671-5411.2013.03.012] [Medline: 24133513]
22. Hindricks G, Taborsky M, Glikson M, Heinrich U, Schumacher B, Katz A, et al. Implant-based multiparameter telemonitoring of patients with heart failure (IN-TIME): a randomised controlled trial. Lancet 2014 Aug 16;384(9943):583-590. [doi: 10.1016/S0140-6736(14)61176-4] [Medline: 25131977]

Abbreviations

CRT: cardiac resynchronization therapy
CRT-D: cardiac resynchronization therapy defibrillator
CRT-P: cardiac resynchronization therapy pacemaker
DOT-HF: Diagnostic Outcome Trial in Heart Failure trial
ICD: implantable cardioverter-defibrillator
CIED: cardiovascular implantable electronic device
CIED^{+}: cardiovascular implantable electronic device with bioimpedance algorithm
CIED^{-}: cardiovascular implantable electronic device without bioimpedance algorithm
IN-TIME: Influence of Home Monitoring on Mortality and Morbidity in Heart Failure Patients With Impaired Left Ventricular Function trial
IQR: interquartile range
HR: hazard ratio
LIMIT-CHF: Lung Impedance Monitoring in Treatment of Chronic Heart Failure trial
OPTILINK-HF: Optimization of Heart Failure Management using OptiVol Fluid Status Monitoring and CareLink trial
SD: standard deviation

Edited by G Eysenbach; submitted 07.08.17; peer-reviewed by I Kedan, P Athilingam, JB Park, K Antypas; comments to author 31.08.17; revised version received 21.09.17; accepted 11.10.17; published 23.11.17.

Please cite as:
Smeets CJP, Vranken J, Van der Auwera J, Verbrugge FH, Mullens W, Dupont M, Grieten L, De Cannière H, Lanssens D, Vandenberk T, Storms V, Thijs IM, Vandervoort PM
Bioimpedance Alerts from Cardiovascular Implantable Electronic Devices: Observational Study of Diagnostic Relevance and Clinical Outcomes
J Med Internet Res 2017;19(11):e393
URL: http://www.jmir.org/2017/11/e393/
doi:10.2196/jmir.8066
PMID:29170147

Original Paper

Development and Validation of a Taxonomy for Characterizing Measurements in Health Self-Quantification

Manal Almalki[1,2*], PhD; Kathleen Gray[1*], PhD; Fernando Martin-Sanchez[1,3*], PhD

[1]Health and Biomedical Informatics Centre, Melbourne Medical School, The University of Melbourne, Melbourne, Australia

[2]Faculty of Public Health and Tropical Medicine, Health Informatics Dept., Jazan University, Jazan, Saudi Arabia

[3]E-Health Research Unit, National Institute of Health Carlos III, Madrid, Spain

[*]all authors contributed equally

Corresponding Author:
Manal Almalki, PhD
Faculty of Public Health and Tropical Medicine
Health Informatics Department
Jazan University
Level 1
Prince Ahmed Bin Abdulaziz Street
Jazan, 82726
Saudi Arabia
Phone: 966 173210584 ext 102
Fax: 966 173210579
Email: manal.almalki1@gmail.com

Abstract

Background: The use of wearable tools for health self-quantification (SQ) introduces new ways of thinking about one's body and about how to achieve desired health outcomes. Measurements from individuals, such as heart rate, respiratory volume, skin temperature, sleep, mood, blood pressure, food consumed, and quality of surrounding air can be acquired, quantified, and aggregated in a holistic way that has never been possible before. However, health SQ still lacks a formal common language or taxonomy for describing these kinds of measurements. Establishing such taxonomy is important because it would enable systematic investigations that are needed to advance in the use of wearable tools in health self-care. For a start, a taxonomy would help to improve the accuracy of database searching when doing systematic reviews and meta-analyses in this field. Overall, more systematic research would contribute to build evidence of sufficient quality to determine whether and how health SQ is a worthwhile health care paradigm.

Objective: The aim of this study was to investigate a sample of SQ tools and services to build and test a taxonomy of measurements in health SQ, titled: the classification of data and activity in self-quantification systems (CDA-SQS).

Methods: Eight health SQ tools and services were selected to be examined: Zeo Sleep Manager, Fitbit Ultra, Fitlinxx Actipressure, MoodPanda, iBGStar, Sensaris Senspod, 23andMe, and uBiome. An open coding analytical approach was used to find all the themes related to the research aim.

Results: This study distinguished three types of measurements in health SQ: body structures and functions, body actions and activities, and around the body.

Conclusions: The CDA-SQS classification should be applicable to align health SQ measurement data from people with many different health objectives, health states, and health conditions. CDA-SQS is a critical contribution to a much more consistent way of studying health SQ.

(J Med Internet Res 2017;19(11):e378) doi:10.2196/jmir.6903

KEYWORDS

health; self-management; self-experimentation; wearables; quantified self; taxonomy; classification

Introduction

People may use wearable tools acquired in direct-to-consumer market places to undertake various processes or activities related to personal health care (eg, data collection and data analysis) as part of self-selected real-life and Web-based communities [1-3]. This practice can be referred to as health self-quantification (SQ) [4].

The use of wearable tools for health SQ is growing in popularity. Globally, the total number of self-quantifiers is unknown, but data available from Quantified Self Meetup groups from Keyhole websites and about the volume of retail sales and shipments of wearable tools indicate that more people are becoming interested in health self-care every year. The Quantified Self international social movement, founded in 2007, has grown up over a decade. In 2007, there was only one group, in California, United States; by August 2017, there were 238 groups worldwide as indicated by the Quantified Self Meetup website. The number of members of these groups increased from around 14,000 in November 2012 to over 83,000 in August 2017.

In social networks, the Keyhole hashtag analytics tool shows that hashtags such as #QS, #wearables, and #smartwatches have a wide reach (reach refers to the number of unique people who may see the posts from Twitter and Instagram). For example, from April 30, 2016 to May 31, 2016, the hashtags of #QS involved over two million people who were talking about health, fitness, weight, health data, body measures, apps, etc as denoted by the Keyhole's word cloud.

The volume of retail sales is increasing because people do not want to limit themselves to the functions that are enabled by mobile phone apps alone. They are striving to take advantage of more advanced features enabled by associated wearable tools because these features claim to enable better maintenance or improvement of their health status [5]. In the United States, the retail sales of these tools increased over the first 8 months of 2015 from US $343.5 million to US $754.8 million [6]. Around the world, Apple sold around 13 million units of smart watches in its first year of sale (ie, April 2015), which doubled the sales of the first iPhone [7].

The volume of shipments of wearable tools is also going up. The number of shipments of wearable tools nearly tripled from 2013 to 2015 from 13 million units in 2013 to around 34 million units in 2015. This level of shipments was expected to expand rapidly over the 5 years to 2020 [8,9]. Another study stated that more than 61 million wearable tools would be shipped to mobile phone users in 2017 [10].

Although health SQ may be done as a solitary, private activity, many people choose to share their experiences, reflections, and data. Web-based platforms such as PatientsLikeMe, Genomera, and CureTogether now cater to self-quantifiers. For example, in April 2015, PatientsLikeMe allowed 38,000 members with multiple sclerosis to link and display activity data from their Fitbit trackers [11]. In addition, by 2018, 70% of health care organizations worldwide were predicted to invest in health technologies including apps and wearable tools, according to IDC Health Insights as cited in [12].

Crowdsourcing these data and integrating them with self-quantifiers' clinical data and history could enable health data analytics research to investigate individuals' health self-care activities and examine their effectiveness [2,13]. Findings from such research could lead to the development of better-personalized health interventions and ultimately could improve health care in ways that benefit the many people who use a variety of tools to engage in different sorts of health SQ [14,15]. However, these population-scale benefits of health SQ cannot be realized fully without a common language for describing the measurement data generated by the use of the various SQ apps and tools [16]. The lack of a common language or taxonomy in this field prevents the systematic investigations that are needed to provide evidence of sufficient quality to determine whether and how health SQ is a worthwhile health care paradigm, and so to advance the use of SQ in health self-care [17].

The aim of this paper, therefore, was to investigate a significant sample of wearable tools and services to build and test a taxonomy of health SQ measurements. We called this the classification of data and activity in self-quantification systems (CDA-SQS).

Methods

This section shows the phases of developing CDA-SQS, as follows.

Phase 1: Building a Preliminary Taxonomy of Health SQ Measurements

This phase was concerned with building a preliminary taxonomy of the measurements that self-quantifiers can take using wearable tools to achieve their health objectives (eg, measuring being active by capturing data that you have walked 10,000 steps per day). To achieve this, first, a sample of health SQ tools and services (N=8) was selected to be examined in mid 2012. Then, the analysis of the measurements that could be generated by these tools was carried out by using an open coding technique (Figure 1, see also [18]). The following two subsections explain these steps.

Selection of Health SQ Tools

This subsection provides information on the selected tools, selection criteria, and tools' descriptions.

A sample of health SQ tools and services (N=8) was selected to be examined in this review study: (1) Zeo Sleep Manager, (2) Fitbit Ultra, (3) Fitlinxx Actipressure, (4) MoodPanda, (5) iBGStar, (6) Sensaris Senspod, (7) 23andMe, and (8) uBiome. The sample needed to reflect the different features related to data collection that were available for health self-quantifiers to use and those that Quantified Self groups identified. These included manual and automatic data collection, single-use data collection, and data types, as shown in Table 1.

Figure 1. Open coding and comparative assessment used to build a taxonomy of health self-quantification (SQ) measurements.

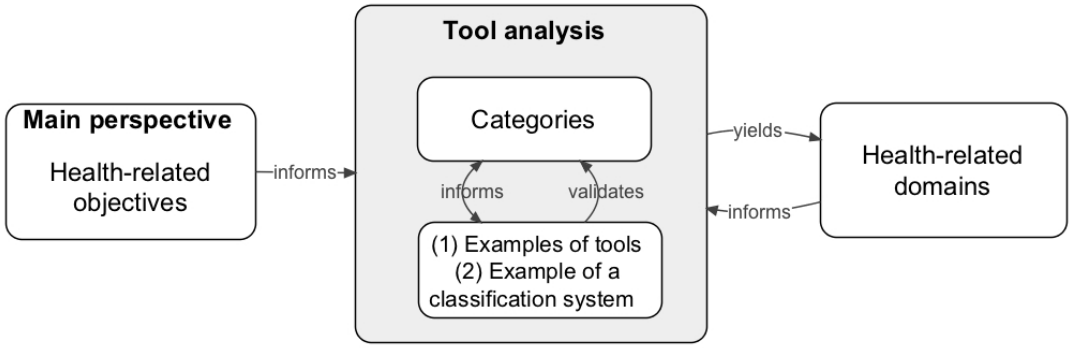


Table 1. Selection criteria of self-quantification (SQ) tools. Table rows do not imply direct correspondences among health SQ tools but rather indicate the presence of a feature in the selected SQ tool or service.

Feature related to data collection	Tool number							
	1	2	3	4	5	6	7	8
Manual data collection			X	X				
Automatic data collection	X	X			X	X		
Single-use data collection							X	X
Data types	Sleep hours and quality	Body movement–related data, for example, steps taken	Blood pressure and pulse	Mood	Blood glucose	Environmental data, for example, ambient humidity	Genome data (single nucleotide polymorphism profile)	Microbiome data

These tools were chosen from a list of forty-two health SQ tools that were being acquired in 2012 to build a SQ research laboratory in the Health and Biomedical Informatics Centre (HaBIC) at the University of Melbourne. The tools acquired for this lab (listed in Multimedia Appendix 1) represented the best selection available in the consumer market place at that time to measure different aspects of health, and so they reflected the existing diversity of health data types that could be generated from using health SQ tools.

The description of the selected tools is based on the following information sources and information-gathering methods:

- Physical inspection of some of the selected tools by researchers. For example, Fitbit Ultra was used by author FM, 23andMe by author KG, and MoodPanda by author MA.
- Information pages that were available on the manufacturer's or service provider's website.
- The manufacturer's technical manuals.
- Articles published in academic journals (eg, [19,20]).
- A number of openly accessible Web-based sources such as the Quantified Self guide; app stores including the Apple Store, Google Play (formerly the Android market), Microsoft Store, and BlackBerry World; and websites such as Vandrico, Wellocracy, and the PatientView directory of health apps.
- Blogs [21].
- Quantified Self "show and tell" meetings; during the PhD study, the researchers were active members of the Melbourne Quantified Self MeetUp group and attended several "show and tell" meetings of this group. During these meetings, Quantified Self members talked about the tools they used for their health SQ. Most of them were coincidentally the same tools that were included in this study (eg,, Zeo Sleep Manager, 23andMe, MoodPanda, and Fitbit Ultra); therefore, based on such users' accounts, the researchers were able to further enhance the content of the tools review.
- Upon requesting permission to use pictures of SQ tools in publications related to our study, a number of commercial representatives asked to review the researchers' description of their product. On the basis of that, they offered additional information and suggestions to improve this description, which served as a double check and validation of this part of the study.

It is worth noting that the description of the sampled tools in this paper is based on versions and sources that were available in 2012; because they were selected and reviewed at that time, they are described below in the past tense. Some technical features may have changed by the time of publication. For example, the 2016 version of Fitbit was called Fitbit AltaHR. This system had no base station to be attached to the user's computer and used Bluetooth to connect the wearable device to the computer or mobile phone. In addition, the Fitbit physical design was changed from a clip that could be attached to clothing into a bracelet that could be worn on the wrist. However, changes in the technical features of the tools that were sampled—and indeed the discontinuation of some tools and the

emergence of other new ones—have not altered the value of the method used in this study to build the taxonomy of health SQ measurement. This is because the focus here is not on these tools' technical features as such but rather on using their description as a first step in the process of typifying the health data types collected by many such tools.

Here are the descriptions of the eight selected tools:

1. Zeo Sleep Manager was a tool for tracking sleep quality, that is, the amount of hours spent in the following four states: light sleep, deep sleep, very deep sleep or dreaming stage, and waking. It comprised a headband that users wore during sleep. The headband measured the brain's electrical signals and provided a quantitative sleep quality value called the Z-score [19]; on the basis of that, it indicated the state that the person is in. These signals were sent to a bedside-clock device to be stored on an attached secure digital memory card and then analyzed. To see a history of sleep states, the user needed to have Internet access and create an account on mysleep.myzeo.com. The company stopped offering their services in 2013 and has not been accessible since then.
2. Fitbit Ultra was a tool for tracking movement including steps taken, stairs climbed, distance travelled, and movement during sleep; calories burned; and hours of sleep. It consisted of a clip that could be attached to clothing; a base station that could be attached to a Windows personal computer or a Mac, which connected the clip with the paired computer or mobile phone; and the Fitbit app for visualizing, analyzing, and sharing the collected data.
3. Fitlinxx Actipressure was a tool for tracking blood pressure and pulse. It comprised an inflatable rubber cuff and a device that had a large digital panel for displaying blood pressure and pulse readings. The user needed to wear the cuff on the upper arm and press the start button, which caused the cuff to inflate automatically and take measurements. To build and review the history of measurements collected, the user needed to have Internet access and create an account on the ActiHealth website where she or he could access the widget ActivePressure and see their blood pressure and pulse readings. This device connected to the user's computer wirelessly via an ActiLink personal access point that plugged into the computer's universal serial bus (USB) port.
4. MoodPanda was a mobile phone–based app for tracking happiness. It allowed users to rate their happiness on a scale from 0 to 10, where 0 was very unhappy and 10 was very happy. Users could also add a brief comment about what was influencing their mood and share their ratings and comments with friends. The app was compatible with different mobile platforms and could be downloaded from app stores such as Apple Store, Google Play (formerly the Android market), Microsoft Store, and BlackBerry World.
5. iBGStar was a blood glucose meter. It consisted of the iBGStar Diabetes Manager app and a blood glucose meter that could be used on its own or attached to an iPhone or iPod touch through the USB port for displaying, tracking, and communicating data of particular interest to people with diabetes. A blood sample was obtained by a lancing tool and placed on the test strip for measuring the blood glucose level. Once the blood glucose meter was attached to the mobile phone, the iBGStar Diabetes Manager app was launched. Then, the readings were automatically logged in the app. If the meter was used alone, the data were saved in the meter's memory and loaded onto the mobile app upon next connecting to the Internet. The app also allowed the user to email these collected readings to others including health care professionals, or transfer them to computers for storing or analysis.
6. Sensaris Senspod was a device that captured environmental data in real time and sent them via Bluetooth to a paired mobile phone. It could be installed in homes, offices, etc to capture ambient noise, humidity, temperature, and carbon monoxide and nitrogen oxide levels. Senspod was provided with an Android application and access to the Web interface. A user could login through their mobile phone or computer to their account on Sensaris.com to read and share the collected data.
7. 23andMe was a Web-based service for performing a genetic test consisting of single nucleotide polymorphisms profile analysis. Users needed first to order a kit from 23andMe in California, United States through the 23andMe website. The kit consisted of a tube where a sample of the users' saliva or a cheek swab was placed. Then, the user needed to create a personal account on 23andMe.com and register the tube identity number before sending it to the 23andMe lab in the United States. Within 4 to 6 weeks, the user received reports which described many of their genes and genetic variants that could be associated with risk of diseases and also provided some information about their ancestors [22]. These reports could be used for predicting diseases that might affect the person in the future and hence, could enable a proactive approach to health management [23]. Furthermore, they could be used for designing more personalized treatment of health conditions [22]. Users could also anonymously compare their own results with others who had genomes like theirs, as well as with the latest scientific research findings.
8. uBiome was a service for analyzing the deoxyribonucleic acid of bacteria that exist in the skin, ears, mouth, sinuses, genitals, and gut. Users needed first to order a kit from uBiome in San Francisco, United States through the website. The kit consisted of strips for taking swabs from different body parts (eg, skin, ears, and mouth). The user received a participant identity number upon ordering the kit; this was needed for signing up to create a personal account on uBiome.com. Next, after taking swabs, the user sent the kit back to the uBiome lab. The user then needed to log in to their account to see and analyze the results. A variety of analysis widgets (eg, percentages and distribution frequencies) and related data viewers (eg, bar chart, pie charts, and logbook) could be accessed via the user's personal account. Users could also anonymously compare their own results with others who had microbiomes like theirs, as well as with the latest scientific research findings.

Analysis of Data Types From Health SQ Tools

An inductive content analysis method—also known as open coding—was employed for analyzing measurements, as shown in Figure 1 [24,25]. During open coding, the tools were reviewed and compared iteratively. Specifically, one researcher examined all the data types (eg, sleep and blood glucose) that could be acquired by the eight tools and inductively assigned them to categories. The data types that were similar were grouped together to create a category or a class. If a data type did not fit with the previously created class, a new class would be added. The classes were then grouped into overarching health-related domains. The coding process was repeated until domains and classes reached a stable state, that is, additional tools did not yield new domains or classes. Open coding was conducted by author MA. Authors KG and FM were involved as reviewers in this step to ensure that each theme was representative, as well as to check how similar they were to each other within a group and how different from codes in every other group. On the basis of this process, the arrangement of codes was refined.

The preliminary taxonomy based on analysis of eight tools was revisited (as shown in Figure 1) in relation to the other tools acquired for the HaBIC health SQ lab. Going through this comparative assessment process confirmed that it would not expand the taxonomy significantly if this study were to extend to in-depth examination of all of these tools. In fact, even this wide range of tools did not yet support many types of health-related measurements with which the researchers were familiar (eg, no tools were available that could measure exposure to all air pollution particulates).

Thus, reaching saturation in terms of describing all possible data types in all health-related domains in this phase of the study was not feasible. So there was a need to ground our preliminary taxonomy in relation to an external rigorous classification system, to build the taxonomy as exhaustively as possible. This was done in phase 2.

Phase 2: Refining the Taxonomy in Relation to Other Classification Systems

The second phase, in mid 2013, sought to refine the preliminary taxonomy (as shown in Figure 1) through comparative assessment with an already established system for classifying health.

For this purpose, we looked for a classification system that met the following criteria: developed by an internationally recognized health organization, comprehensive, health and functioning must be the basic organizing concept, able to fit within an external framework to contribute to a more consistent way of studying health SQ, and able to accommodate conventional and unconventional observations of potential influences on an individual's health. Three international classification systems were identified and considered: the World Health Organization's International Classification of Functioning, Disability, and Health (WHO-ICF) [26]; the European Directory of Health Apps [27]; and Happtique [28]. Comparative descriptions of each one are provided in Multimedia Appendix 2.

WHO-ICF was selected because it met all the stated criteria, and it provided the most comprehensive list of health measurements—it had more than 1400 categories [29], whereas the European Directory of Health Apps and Happtique, respectively, had around 60 and 300 categories. Health and functioning were its basic organizing concept [26,30]. It could be coupled with external frameworks for describing health concepts [26,30]. It defined the individual's health function or disability in relation to what they can do or cannot do as a dynamic interaction between the individual, their personal factors (eg, age and gender), and their environmental factors (eg, ambient weather temperature and humidity and use of communication technologies such as mobile phones) [26,31,32].

At the end of phase 2, the taxonomy was expanded through this comparison and refinement of categories; however, it remained theoretical. So a third phase of the study was carried out to test it in relation to real-world health SQ.

Phase 3: Validating the Taxonomy in Relation to Self-Quantifiers' Practices

To validate the taxonomy, it was examined in relation to self-reported activities of health self-quantifiers. This information was derived from the results of an international Web-based survey that we conducted in 2014 to explore many aspects of people's use of a wide range of health SQ tools and services. More details about the broad aim, design, and results of this survey is available in [33].

Pertinent to validating the theoretical taxonomy of health SQ measurements, among the survey's 67 questions were multiple-choice questions, as well as open-ended questions to elicit information about the names and the number of tools used by an individual and the kinds of data that the individual collected using these tools. The survey results enabled us to revisit our categories in relation to responses about the use of 130 different SQ tools. These responses were provided by 103 self-quantifiers who varied in age, gender, health status, and motivation.

For the purpose of validation, the researchers qualitatively analyzed these responses using deductive thematic coding based on the theoretical taxonomy. Once again, initial coding was conducted by author MA and reviewed by authors KG and FM in this step. We observed whether and to what extent it was possible to account for the tools used and the data types generated from these tools within corresponding health domains and categories in the taxonomy.

Results

Phase 1 led to the development of a preliminary taxonomy that we called CDA-SQS, as shown in Table 2. Phase 2 led to the development of a model of the measurements in health SQ, as shown in Figure 2, as well as to the improvement of the taxonomy in terms of increasing the number of health measurement categories that it could account for, as illustrated in Table 3. Phase 3 led to refinement to the CDA-SQS model and showed the applicability of CDA-SQS for characterizing measurements in health SQ. The following paragraphs explain the results from these three phases in more detail.

Table 2. The health domains and corresponding categories of the selected self-quantification (SQ) tools. The numbers from 1 to 8 indicate the SQ tool or service we analyzed. "X" indicates the SQ tool or service's functionality to capture this type of measurement.

Domain name	CDA-SQS[a]		SQS							
Health domain	Category	Subcategory	1	2	3	4	5	6	7	8
Body functions	Mental functions	Sleep	X	X						
	Mental functions	Emotions				X				
	Cardiovascular system	Blood pressure			X					
	Endocrine system	Blood glucose					X			
Body structures	Cell structure	Genes, deoxyribonucleic acid, etc							X	
	Microbial structure in skin, gut, etc	Names, number, types, etc								X
Body actions and activities	Mobility	Walking		X						
Around the body	Natural environment	Climate or weather						X		

[a]CDA-SQS: Classification of data and activity in self-quantification systems.

Figure 2. The interactive model of classification of data and activity in self-quantification systems (CDA-SQS).

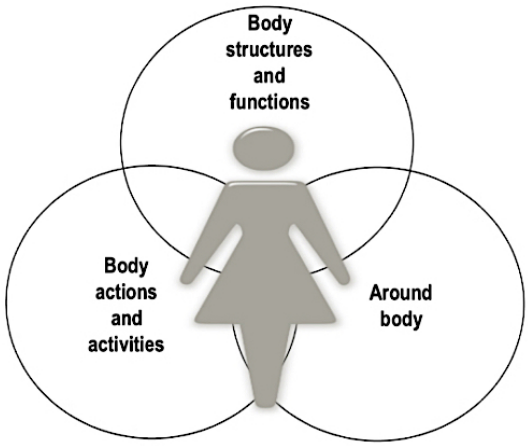


Table 3. A two-level summary of the classification of data and activity in self-quantification systems (CDA-SQS). Table rows do not imply direct correspondences among health self-quantification measurements but rather indicate the wide variety of combinations that a self-quantifier may choose to explore.

Body structures	Body functions	Body actions and activities	Around the body
Cell structure	Mental functions	Learning and applying knowledge	Natural and built environment
Microbial structure in skin, ears, mouth, sinuses, genitals, and gut	Sensory functions	General tasks and demands	Relationships and attitudes
Structure of brain	Sensation of pain	Communication	Community, social, and civic life
The eyes, ears, nose, mouth, pharynx, and larynx	Voice and speech	Mobility	Major life areas
Structure of cardiovascular system	Functions of the cardiovascular system	Self-care	Products and technologies
Structure of immune system	Functions of the hematological system	Domestic life	Time
Structure of respiratory system	Functions of the immune system	Recreation and leisure	
Structure of digestive system	Functions of the respiratory system	Religion and spirituality	
Structure of endocrine glands	Functions of the digestive system		
Structure of urinary system	Functions of the metabolic system		
Structure of reproductive system and pelvic floor	Functions of the endocrine system		
Structure of musculature	Genitourinary functions		
Structure of spinal cord	Reproductive functions		
Structure of skin, hair, and nails	Neuromusculoskeletal and movement-related functions		
	Functions of skin, hair, and nails		

Results From Phase One

Table 2 presents the results from the open coding that was conducted in the first phase. It shows the categories of what people would be measuring, if they were using the SQ tools that we examined, to understand their body and health status. These tools were grouped by the data types into (1) inside the human body, in which data were related to bodily functions (N=4) and structures (N=2); (2) on the body, which was concerned with bodily actions and activities (N=1); and (3) around the body (N=1), which pertained to bodily surrounds including environmental aspects that might affect the individual's health. Therefore, CDA-SQS categorizes measurements into three main health-related domains: body structures and functions, body actions and activities, and around the body.

Results From Phase Two

Table 3 illustrates the CDA-SQS domains and related categories; this represents the result from the second phase. The CDA-SQS presents the three domains and their related categories and subcategories in a hierarchical way. The following paragraphs describe CDA-SQS through comparing and contrasting it with WHO-ICF. Multimedia Appendix 3 displays a detailed description of the taxonomy developed from this study.

The WHO-ICF taxonomy categorizes measurements into three main health-related domains: body functions and structure, activity and participation, and environmental factors (eg, the natural and built environment). In CDA-SQS, the domain "in body" is more or less similar to the corresponding one in WHO-ICF because they both describe aspects related to functioning at the level of the body [34]. However, CDA-SQS added new categories. For example, the class "microbial structure in skin, ears, mouth, sinuses, genitals, and gut" was added to the category "body structures."

In WHO-ICF, the second domain is about aspects of functioning at the level of the individual (ie, one's activities) and at the social level (ie, one's participation in society), whereas environmental factors are conceived as facilitators or barriers to the person's functioning at these levels [34]. However, the distinction between functioning at the level of the individual and functioning at the social level is not clear; it has proved problematic to operationalize, and it has been argued that categories related to the activity and participation domain should belong to two different domains [31]. This was done in CDA-SQS as follows.

In CDA-SQS, activities now belong to the "body actions and activities" domain because they are related to functioning at the level of the individual, whereas participation and environmental factors belong to the "around the body" domain because they describe surrounding aspects that may affect the individual's health. Consequently, the categories related to the domains "activity and participation" and "environmental factors" in WHO-ICF were rearranged to fit the purpose of CDA-SQS. For example, categories such as "mobility," "general tasks and demands," "communication," "learning and applying knowledge," and "self-care" were placed under the body actions and activities domain, whereas "major life areas" and "community, social, and civic life" were placed under the "around the body" domain.

Some categories in WHO-ICF were complicated, such as "functions of the cardiovascular, hematological, immune, and respiratory systems." Such categories were rendered into a number of corresponding categories; in this instance, it was rendered into four categories: functions of the cardiovascular system, functions of the hematological system, functions of the immune system, and functions of the respiratory system.

In contrast, some WHO-ICF categories such as "interpersonal interactions and relationships" and "attitude" created distinctions that were not useful. These were combined into one CDA-SQS category, in this case called "relationships and attitudes." We argue that for the purpose of health SQ, measuring interactions and relationships is more than just counting the number of them. What most self-quantifiers are interested to track and explore are the attitudes that they hold toward these social relations, the attitudes they experience within them, and how these attitudes affect their personal well-being.

Overall, CDA-SQS is consistent with WHO-ICF in major respects. It structures domains into categories and subcategories, as WHO-ICF does. It can fit within external frameworks to capture contextual data about self-quantifiers for studying their health status, as WHO-ICF can. For example, CDA-SQS can be integrated with a framework called minimal information about SQ experiment [35,36]; hence, data about a self-quantifier's personal factors (eg, age, gender, and education level) and more can be captured.

Results From Phase Three

Our survey showed that to better understand the health status, individual self-quantifiers tend to collect multiple types of health data, which may be in different health domains, for example, the connection between sleep patterns and weather patterns. Therefore, the CDA-SQS model is illustrated as three intersecting circles, as depicted in Figure 2. This refinement is in agreement with WHO-ICF; both systems conceive of health domains and categories as interactive and an individual's health as the product of interaction among components.

The CDA-SQS classification was also found to be applicable to all of different kinds of health measurements (N=130) that survey participants reported taking with SQ tools. Every type of measurement could be fitted into one of these health-related domains: "body structures and functions" (N=53), "body actions and activities" (N=39), and "around the body" (N=16). Each of these domains has a hierarchy of categories and subcategories. Multimedia Appendix 4 shows the classification of a subset of kinds of measurements taken using SQ tools (N=108) within these three domains. The remaining 22 tools were found to capture a range of health aspects that belong to more than one health domain denoted by the CDA-SQS; for the sake of clarity, they were listed in Multimedia Appendix 4 but not presented in the appendix table.

Discussion

CDA-SQS provides a common language for describing measurements generated from SQ tools. This was a largely untapped area in prior literature: our systematic literature reviews [4,37] revealed limited attempts from researchers to classify SQ data in a comprehensive or systematic manner. One study collected 209 users' reviews from the Quantified Self website [38] and, using inductive content analysis, classified health SQ data into five categories: body state (eg, physical and physiological), psychological state and traits, activities (eg, exercise, eating, and sleeping), social interactions, and environmental and property states. However, we found this classification inadequate to account for all of the health-related aspects in each category that our research uncovered. CDA-SQS provides a comprehensive description of measurements generated from SQ tools in comparison with the classification developed by [38].

CDA-SQS is also more versatile than the non-peer reviewed classifications we identified. For example, CDA-SQS identifies more health categories than the myhealthapps directory 2015-2016 [39]. Although that directory was built by inductively analyzing 300 health-related apps recommended by consumers for health self-care, it classifies these apps into only two main domains (ie, disability and health) and 21 corresponding categories. In contrast, CDA-SQS was developed initially by inductively analyzing eight SQ tools and nominating three main health domains; however, the subsequent phases in its development process led to generating 43 categories under these three domains, further broken down into 124 subcategories.

In CDA-SQS, health and functioning is the basic organizing concept; this is in line with the current orientation in delivering health care [30,40]. Furthermore, it can be used to augment other analytical tools that are necessary to investigate health SQ rigorously: it fits with our SQS taxonomy for classifying tools and services [41]. It can be integrated with our theoretical framework for dissecting health SQ activity [37]. It is compatible with models such as our Personal Health Information Self-Quantification Systems [33] and our health SQ chain value model [4]. In this way, CDA-SQS is a critical contribution to a more holistic and consistent way of studying health SQ; SQ data can be put into an individual's contexts of use—in relation to the measurements denoted by CDA-SQS, to the tools and services described by SQS taxonomy, to the health self-management activities illustrated by our models, and to the personal demographics and the community connections shown by the health SQ activity framework. Furthermore, CDA-SQS can account for relatively conventional observations (eg, cardiovascular function) and also less conventional observations (eg, gut microbes and environmental exposures) about an individual's health. This in turn may help in studying the interplay of these aspects of health and ultimately in developing more personalized health interventions [42].

This study has the strength that arises from collaborative use of mixed qualitative methods within a research team. Critically, it used both inductive and deductive logic to examine the health-related measurements generated from the use of SQ tools and services. This led not only to developing a classification schema but also to establishing a conceptual framework for describing measurements in health SQ. However, this study has some limitations. In the inductive content analysis, complete interrater reliability analysis could not be conducted because this study was conducted as part of author MA's PhD student, and so MA had to be the only coder in phases 1 and 3 [43]. As explained in the Methods section, the other authors were MA's PhD supervisors and so could provide oversight and critique but not full reliability testing. Thus, a follow-up study to test CDA-SQS performance on a wide range of health SQS in laboratory or field trials would shed more light on the results of this study. Other methods too could be used to test the validity of phase 2, for example, eliciting review and feedback from health care professionals and from more members of the Quantified Self community.

This study has articulated a formal methodological approach for describing and distinguishing the different types of measurements in health SQ. The resulting CDA-SQS classification is applicable to many different types of scientific studies that involve SQ measurements taken by people regardless of their health objective, health status, or health condition, and regardless of the tool or service they use for this purpose.

Establishing protocols that can be used to work with health self-quantifiers' measurements is a necessary biomedical informatics foundation for reporting experimental results in this area. CDA-SQS has shown its usefulness in establishing a common format and reporting guideline for health SQ studies [35,36]. These informatics tools may serve as the basis for improving the design of research databases, including public repositories, where users of health SQS choose to contribute their data. Our future studies aim to test CDA-SQS applicability for such purposes.

Acknowledgments

This study was supported in part through a PhD scholarship from the University of Melbourne, a PhD scholarship from Jazan University, and a seed funding grant and top-up scholarship from the Melbourne Networked Society Institute.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of SQ tools available in the HaBIC SQ lab.

[PDF File (Adobe PDF File), 37KB - jmir_v19i11e378_app1.pdf]

Multimedia Appendix 2

Description of and comparison between three classification systems.

[PDF File (Adobe PDF File), 42KB - jmir_v19i11e378_app2.pdf]

Multimedia Appendix 3

Classification of data and activity in self-quantification systems.

[PDF File (Adobe PDF File), 49KB - jmir_v19i11e378_app3.pdf]

Multimedia Appendix 4

Applicability of CDA-SQS.

[PDF File (Adobe PDF File), 42KB - jmir_v19i11e378_app4.pdf]

References

1. Lupton D, Jutel A. 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. Soc Sci Med 2015 May;133:128-135. [doi: 10.1016/j.socscimed.2015.04.004]
2. Lupton D. Digitised health, medicine and risk. Health Risk Soc 2016 Mar 04;17(7-8):473-476. [doi: 10.1080/13698575.2015.1143247]
3. Swan M. Health 2050: the realization of personalized medicine through crowdsourcing, the Quantified Self, and the participatory biocitizen. J Pers Med 2012 Sep 12;2(3):93-118 [FREE Full text] [doi: 10.3390/jpm2030093] [Medline: 25562203]
4. Almalki M, Gray K, Martin-Sanchez FJ. Refining the concepts of self-quantification needed for health self-management: a thematic literature review. Methods Inf Med 2016 Aug 15;56(1):46-54. [doi: 10.3414/me15-02-0007] [Medline: 27523820]
5. Hollandlogisticslibrary. 2014. Racing toward a complete digital lifestyle: Digital consumers crave more URL: https://www.hollandlogisticslibrary.com/outbound/resource/racing-toward-a-complete-digital-lifestyle-digital-consumers-crave-more [accessed 2016-10-27] [WebCite Cache ID 6lZtyWZ3K]
6. NPD. 2015. Ownership of activity trackers, smartwatches expected to jump at least 4 percent this holiday season Press release URL: https://www.npd.com/wps/portal/npd/us/news/press-releases/2015/ownership-of-activity-trackers-smartwatches-expected-to-jump-at-least-4-percent-this-holiday-season-according-to-npd [WebCite Cache ID 6ubkGNgdR]
7. Wakabayashi D. WSJ. Apple's watch outpaced the iPhone in first year 2016 URL: http://www.wsj.com/articles/apple-watch-with-sizable-sales-cant-shake-its-critics-1461524901 [accessed 2016-10-27] [WebCite Cache ID 6lZuJVnYR]
8. Statista. 2016. Worldwide shipments of healthcare wearables 2013-2015 URL: https://www.statista.com/statistics/302725/healthcare-wearables-shipments-worldwide/ [accessed 2016-10-27] [WebCite Cache ID 6lZuTnr1z]
9. Statista. 2016. Number of wearable tools worldwide by region 2015 and 2020 URL: https://www.statista.com/statistics/490231/wearable-devices-worldwide-by-region/ [accessed 2016-10-27] [WebCite Cache ID 6lZud2GJ8]
10. Research2guidance. 2017. The market for c will grow to $5.6 Billion by 2017 URL: http://research2guidance.com/the-market-for-mobile-health-sensors-will-grow-to-5-6-billion-by-2017 [accessed 2016-10-27] [WebCite Cache ID 6lZupakB9]
11. PatientsLikeMe. 2015. Study monitors walking activity in people with multiple sclerosis Press release URL: http://news.patientslikeme.com/press-release/patientslikeme-study-monitors-walking-activity-people-ms [accessed 2016-10-27] [WebCite Cache ID 6lZvFraCj]
12. emedcert. 2015. The numbers don?t lie: Wearables are the future of healthcare URL: https://emedcert.com/blog/wearables-statistics-future-of-healthcare [accessed 2016-10-27] [WebCite Cache ID 6lZvTkcM2]
13. Vesnic-Alujevic L, Breitegger M, Guimarães Pereira Â. 'Do-It-Yourself' healthcare? Quality of health and healthcare through wearable sensors. Sci Eng Ethics 2016 Mar 30:- (forthcoming). [doi: 10.1007/s11948-016-9771-4] [Medline: 27029478]
14. Sharon T. Self-tracking for health and the quantified self: re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. Philos Technol 2016 Apr 18;30(1):93-121. [doi: 10.1007/s13347-016-0215-5]
15. Vogt H, Hofmann B, Getz L. The new holism: P4 systems medicine and the medicalization of health and life itself. Med Health Care Philos 2016 Jun;19(2):307-323 [FREE Full text] [doi: 10.1007/s11019-016-9683-8] [Medline: 26821201]

16. Almalki M, Gray K, Martin-Sanchez F. Classification of data and activities in self-quantification systems. 2014 Presented at: HISA Big Data in Biomedicine and Healthcare Conference; April 2014; Australia, Melbourne.
17. Gray K, Martin-Sanchez FJ, Lopez-Campos GH, Almalki M, Merolli M. Person-generated data in self-quantification: a health informatics research program. Methods Inf Med 2017 Jan 09;56(1):40-45. [doi: 10.3414/ME15-02-0006] [Medline: 27782291]
18. Ohlin F, Olsson C. Intelligent computing in personal informatics: Key design considerations. 2015 Presented at: IUI '15 Proceedings of the 20th International Conference on Intelligent User Interfaces; January 2015; Sonoma, United States. [doi: 10.1145/2678025.2701378]
19. Chang J. Self-tracking for distinguishing evidence-based protocols in optimizing human performance and treating chronic illness. 2012 Presented at: AAAI Spring Symposium on Self-Tracking and Collective Intelligence for Personal Wellness; 2012; Palo Alto, United States URL: http://aaai.org/ocs/index.php/SSS/SSS12/paper/view/4323/4670
20. Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. 2010 Presented at: CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2010; Atlanta, United States. [doi: 10.1145/1753326.1753409]
21. Johnfass. 2012. BodyTrack/Fluxtream URL: http://johnfass.wordpress.com/2012/09/06/bodytrackfluxtream [WebCite Cache ID 6lZvhqX6c]
22. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. Nature 2009 Oct 8;461(7265):724-726. [doi: 10.1038/461724a]
23. Carlson B. Can online genetic testing predict the future? Biotechnol Healthc 2008;5(3):11-12 [FREE Full text]
24. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2007;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x]
25. Howitt D. Introduction to qualitative methods in psychology. Harlow, England: Financial Times Prentice Hall; 2010.
26. International classification of functioning, disability and health: ICF short version. Geneva: World Health Organisation; 2001.
27. patient-view.: Patient View; 2012. The European Directory of Health Apps 2012-2013 URL: http://www.patient-view.com/-bull-directories.html [WebCite Cache ID 6ZxDT8uqe]
28. Happtique. 2011. Mobile health has taken the world by storm URL: http://www.happtique.com/wp-content/uploads/HAPP_booklet010212hi.pdf [accessed 2016-10-27] [WebCite Cache ID 6lZvveEdA]
29. Dreinhöfer K. In: Dreinhöfer K, Dieppe P, Günther K, Puhl W, editors. EUROHIP: Health technology assessment of hip arthroplasty in Europe. Berlin, Heidelberg: Springer-Verlag; 2009.
30. Jette A. ptjournal. Toward a common language for function, disability, and health URL: http://ptjournal.apta.org/content/86/5/726.abstract [accessed 2017-10-30] [WebCite Cache ID 6ubmjT1Fm]
31. Badley EM. Enhancing the conceptual clarity of the activity and participation components of the International Classification of Functioning, Disability, and Health. Soc Sci Med 2008 Jun;66(11):2335-2345. [doi: 10.1016/j.socscimed.2008.01.026] [Medline: 18314239]
32. WHO.: World Health Organisation; 2013. How to use the ICF a practical manual for using the International Classification of Functioning, Disability and Health (ICF): Exposure draft for comment URL: http://www.who.int/classifications/drafticfpracticalmanual.pdf [accessed 2017-10-30] [WebCite Cache ID 6ubnxXGxx]
33. Almalki M, Sanchez FM, Gray K. Quantifying the activities of self-quantifiers: management of data, time and health. Stud Health Technol Inform 2015;216:333-337. [Medline: 26262066]
34. CDC. The ICF: An overview URL: https://www.cdc.gov/nchs/data/icd/icfoverview_finalforwho10sept.pdf [accessed 2017-10-30] [WebCite Cache ID 6ubnF5FI6]
35. Lopez-Campos G, Almalki M, Martin-Sanchez F. knowledge.amia. 2014. Towards a representation format for sharable self-monitoring data URL: https://knowledge.amia.org/56638-amia-1.1540970/t-005-1.1543914/f-005-1.1543915/a-429-1.1544384/an-429-1.1544385?qr=1 [accessed 2017-10-30] [WebCite Cache ID 6ubnLF1lW]
36. Lopez-Campos G, Almalki M, Martin-Sanchez F. Proposal for a standardised reporting guideline to annotate health-related self-quantification experiments. Stud Health Technol Inform 2014;202:79-82. [Medline: 25000020]
37. Almalki M, Gray K, Martin-Sanchez F. Activity theory as a theoretical framework for health self-quantification: a systematic review of empirical studies. J Med Internet Res 2016 May 27;18(5):e131. [doi: 10.2196/jmir.5000] [Medline: 27234343]
38. Oh J, Lee U. Exploring UX issues in quantified self technologies. 2015 Presented at: 8th International Conference on Mobile Computing and Ubiquitous Networking; 2015; Hakodate, Japan. [doi: 10.1109/ICMU.2015.7061028]
39. patient-view.: Patient View; 2015. The myhealthapps directory 2015-2016 URL: http://www.patient-view.com/-bull-directories.html [WebCite Cache ID 6ZxDT8uqe]
40. Lines LM, Lepore M, Wiener JM. Patient-centered, person-centered, and person-directed care: they are not the same. Med Care 2015 Jul;53(7):561-563. [doi: 10.1097/MLR.0000000000000387] [Medline: 26067878]
41. Almalki M, Gray K, Sanchez FM. The use of self-quantification systems for personal health information: big data management activities and prospects. Health Inf Sci Syst 2015;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S1 [FREE Full text] [doi: 10.1186/2047-2501-3-S1-S1] [Medline: 26019809]

42. Martin Sanchez F, Gray K, Bellazzi R, Lopez-Campos G. Exposome informatics: considerations for the design of future biomedical research information systems. J Am Med Inform Assoc 2014 May 01;21(3):386-390. [doi: 10.1136/amiajnl-2013-001772] [Medline: 24186958]
43. Armstrong D, Gosling A, Weinman J, Marteau T. The place of inter-rater reliability in qualitative research: an empirical study. Sociology 1997 Aug;31(3):597-606. [doi: 10.1177/0038038597031003015]

Abbreviations

CDA-SQS: classification of data and activity in self-quantification systems
HaBIC: Health and Biomedical Informatics Centre
SQ: self-quantification
WHO-ICF: World Health Organization's International Classification of Functioning, Disability, and Health

Edited by G Eysenbach; submitted 27.10.16; peer-reviewed by Y Lau, R Valdez, C Allen, J Apolinário-Hagen, K Ng; comments to author 15.01.17; revised version received 14.03.17; accepted 30.08.17; published 03.11.17.

Please cite as:
Almalki M, Gray K, Martin-Sanchez F
Development and Validation of a Taxonomy for Characterizing Measurements in Health Self-Quantification
J Med Internet Res 2017;19(11):e378
URL: http://www.jmir.org/2017/11/e378/
doi:10.2196/jmir.6903
PMID:29101092

Review

Patient Health Record Systems Scope and Functionalities: Literature Review and Future Directions

Lina Bouayad[1,2*], PhD; Anna Ialynytchev[2*], PhD; Balaji Padmanabhan[3*], PhD

[1]Department of Information Systems and Business Analytics, Florida International University, Miami, FL, United States

[2]Health Services Research and Development Service, Center of Innovation on Disability and Rehabilitation Research, Tampa, FL, United States

[3]Department of Information Systems and Decision Sciences, University of South Florida, Tampa, FL, United States

[*]all authors contributed equally

Corresponding Author:
Lina Bouayad, PhD
Department of Information Systems and Business Analytics
Florida International University
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 3053485437
Email: lbouayad@fiu.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2019/9/e15796

Abstract

Background: A new generation of user-centric information systems is emerging in health care as patient health record (PHR) systems. These systems create a platform supporting the new vision of health services that empowers patients and enables patient-provider communication, with the goal of improving health outcomes and reducing costs. This evolution has generated new sets of data and capabilities, providing opportunities and challenges at the user, system, and industry levels.

Objective: The objective of our study was to assess PHR data types and functionalities through a review of the literature to inform the health care informatics community, and to provide recommendations for PHR design, research, and practice.

Methods: We conducted a review of the literature to assess PHR data types and functionalities. We searched PubMed, Embase, and MEDLINE databases from 1966 to 2015 for studies of PHRs, resulting in 1822 articles, from which we selected a total of 106 articles for a detailed review of PHR data content.

Results: We present several key findings related to the scope and functionalities in PHR systems. We also present a functional taxonomy and chronological analysis of PHR data types and functionalities, to improve understanding and provide insights for future directions. Functional taxonomy analysis of the extracted data revealed the presence of new PHR data sources such as tracking devices and data types such as time-series data. Chronological data analysis showed an evolution of PHR system functionalities over time, from simple data access to data modification and, more recently, automated assessment, prediction, and recommendation.

Conclusions: Efforts are needed to improve (1) PHR data quality through patient-centered user interface design and standardized patient-generated data guidelines, (2) data integrity through consolidation of various types and sources, (3) PHR functionality through application of new data analytics methods, and (4) metrics to evaluate clinical outcomes associated with automated PHR system use, and costs associated with PHR data storage and analytics.

(J Med Internet Res 2017;19(11):e388) doi:10.2196/jmir.8073

KEYWORDS

personal health record systems; health records, personal; electronic health records; data analytics; medical informatics; patient-centered care; review; health platforms; multiorganizational systems; ultralarge systems

Introduction

The idea of patient health records (PHRs) emerged in the early 1970s [1,2] with the goal of increasing patient engagement and empowerment, which in turn was intended to enable continuity of care, error reduction [3], treatment choice, and patient-provider partnership building [1,2].

An extension of traditional electronic health records (EHRs), PHRs created a patient-centric platform supporting the new vision of health services that enables patient-provider information sharing and collaboration, with the goal of improving health outcomes and reducing costs. In recent decades, great strides have been made toward achieving these far-reaching goals in research and practice. Through the implementation in the United States of the Health Information Technology for Economic and Clinical Health (HITECH) Act passed in 2009, the use of PHR data is becoming more commonplace [4]. As defined by the program, the initial stage of meaningful use encourages providers to integrate technology into medical practice, making vast amounts of patient data available electronically. Later stages of the program focus on empowering patients by providing them with online access to their heath data.

The use of PHRs has grown since the rise of mobile computing and advancement of patients' technical aptitude. As an extension of EHRs, PHRs have been developed to enable patients to manage their own health care. These records include (1) EHR-transmitted data such as laboratory results and summary of care, and (2) patient-generated data such as symptoms. The amount of overlap in terms of data and functionalities between the EHR and PHR depends on the type of implementation: tethered, interconnected, or stand-alone [5]. Functionalities available through the PHR are intended to be used by patients, rather than by providers, and include appointment scheduling, prescription refill, and secure messaging [6]. The newly developed PHRs created a complementary source of clinical data such as patient-reported outcomes [7-9], physician ratings [10], medication adherence [11], and social support [12,13], and they allow for new data analytics techniques to detect, measure, and predict health-related outcomes. The United States has been a leader in the field of PHR data analytics. One reason for the growth of health care analytics in the United States is the incentivization of such research through federal initiatives to deliver patient-centered care and quality-driven payment models [14,15]. The Partnership for the Future of Medicare [15] states that innovative methods, such as email consultations and self-monitoring, must be used to achieve individualized, effective care. Additionally, Medicare strives to make health care data more readily available and accessible, including quality and performance metrics. Taken together, these initiatives support health care data collection and utilization in the United States, making PHR analytics more feasible. However, the full potential of PHR cannot be realized until we have a better understanding of PHR data content, formats, and sources.

Tremendous amounts of patient data are now available through PHR systems. With patients' permission, these data, along with the application of advanced data mining and machine learning, can provide significant new opportunities in research. For instance, models in areas such as disease prediction, patient risk assessment, and early symptom detection can now be improved, leading to major advances in health outcomes and cost optimization. However, along with new opportunities provided by PHR systems come data and user-related challenges. Data-related issues such as quality, privacy, and security pertain to collection, safe storage, and processing of large quantities of patient data from distributed information systems. Also, patients previously excluded from access to such systems may lack the expertise to understand the data [16].

This review assessed the scope of data and functionalities in PHR systems with the goal of understanding how these affect research on health information systems. The platforms today lack a global standard and vary widely in terms of functionalities, goals, privacy issues, and legal frameworks. Hence, looking at the evolution of PHR data elements through a literature review of US studies, we also investigated opportunities and challenges associated with this emerging platform. While our review and implications are US centric, many of the broader research ideas have emerged from global applications.

Methods

We conducted a review of US literature published from 1950 through 2015 to assess the scope and functionalities available through the PHR, along with associated data elements, formats, and sources. We summarized the results and classified the data content through functional categorization and chronological analysis, and identified gaps in the literature. Based on our findings, we present recommendations for health information systems research.

Eligibility Criteria

In this review, we defined PHR as an electronic record designed for patients to self-manage care [6]. Thus, we focused on data that were either entered by or transmitted to the patient to enable self-care management, regardless of PHR type or brand.

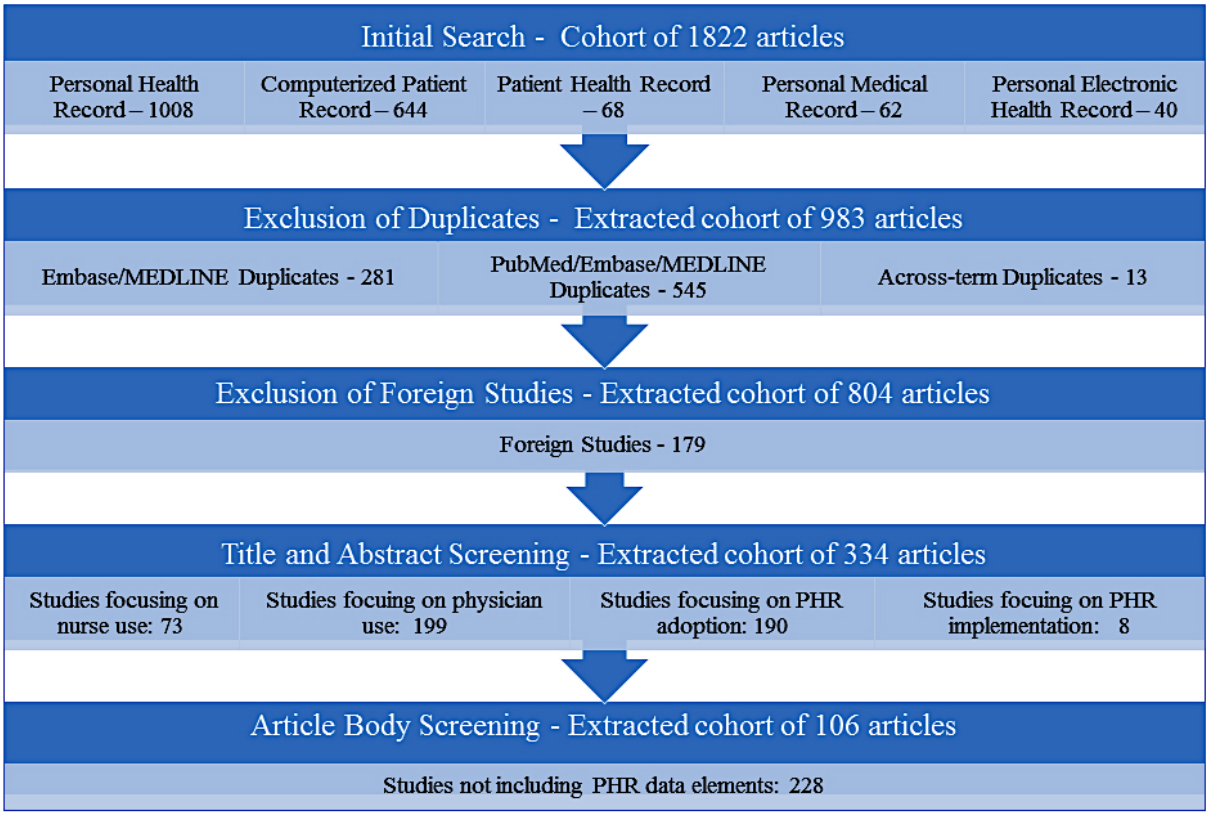
We considered US studies from 1950 through 2015. We limited our search to US-based studies because of variation in ontologies and legal and privacy frameworks across countries. Because we were interested in specific data content available in the PHR, rather than patients' extent of system use, we excluded articles focusing on PHR adoption. Furthermore, we excluded articles containing data intended to be used only by health care providers, and that not to be viewed by patients. For example, articles reporting on physician use of patients' hormone levels to assess risk factors and clinical outcomes were excluded from the analysis because this information was not intended to be used or viewed by patients. Finally, after reviewing the body of articles selected based on title and abstract, we excluded articles that focused on general concepts and did not mention specific data elements present in the PHR.

Data Sources and Search Strategy

To conduct our review, and using used PubMed's Medical Subject Headings (MeSH) database as our starting point, we identified 5 search phrases referring to the PHR: (1) personal

health record, (2) personal medical record, (3) patient health record, (4) computerized patient record, and (5) personal electronic health record. A search of eligible US studies on PubMed from 1950 to 2015 and on Embase and MEDLINE from 1966 to 2015 using the previously defined phrases resulted in 1822 articles (Figure 1). The search results comprised articles containing any of the search phrases in all fields including titles and abstracts.

Figure 1. Literature review results. PHR: patient health record.



Title and abstract screening based on the inclusion and exclusion criteria by 2 reviewers (authors LB and AI) resulted in consideration of 334 articles. Data elements, associated data sources, and analytics techniques were described. The reviewers met after screening every 20 articles to compare results and adjudicate. Consensus was reached regarding (1) the final list of articles to be considered for full-text screening and (2) information extracted from the selected articles.

The full body screening resulted in a total of 106 articles used for data element extraction. Whenever available, reviewers LB and AI recorded the following information in an Excel 2010 (Microsoft Corporation) spreadsheet from each article reviewed in this study: (1) title, (2) author(s), (3) year of publication, (4) PHR data element(s) (ie, data collected by or shared via the PHR), (5) data type(s) (character, number, string, etc), (6) platform(s) (website, app, etc), (7) data storage (Excel database, Oracle, etc), (8) data entry (manual or electronic), (9) source, (10) receiver, (11) details regarding patient use, (12) barriers and issues, and (13) benefits.

Data Categorization

A list of all data elements extracted from the 106 selected articles was further grouped by the reviewers into major data categories. The data categories were based on a taxonomy created in a PHR systematic review published in 2011 by Archer et al [6], which served as a foundation for this work. Categories found in our review but not included in Archer et al's review were identified and validated by a group of clinical informatics experts.

We categorized the PHR data and refined them after consultation with an informal focus group of clinicians. In cases where different terms referred to the same data element (eg, medications, pills, and drugs), we chose 1 of the terms and grouped all synonymous data elements together under this term. Metadata pertaining to PHR functionalities were extracted from the articles and categorized based on content. For instance, articles mentioning the PHR reminder functionality were listed as references for data elements such as appointment reminders and prescription reminders, and were categorized under scheduling and treatments, respectively, as opposed to grouping all reminders under an umbrella "reminders" category. Additionally, some of the PHR data elements could have been included in different categories, depending on the user's perspective. For example, the data elements described as prevention adherence could be viewed by the patient as part of a prevention plan but perceived by the provider as compliance with recommended health procedures and activities. We refined and ordered data categories listed in the results table based on their typical sequence of patient health care delivery. For

example, scheduling data were listed before treatment data, which were listed before outcomes.

Functional Taxonomy and Chronological Analysis

Following PHR data extraction and categorization, we performed a cross-categorical analysis of the data by percentage, source, and format. Additionally, we completed a longitudinal analysis of the time of first mention of the data element in the literature.

Results

Extraction Results

The literature review identified 13 major categories of PHR (Multimedia Appendix 1 [17-117]). At least one data element was included within each of the main categories, and details on the data elements and their corresponding references are provided. In addition to the data elements previously reported in Archer et al's systematic review, this research identified 22 new data elements. Additionally, we distinguished 3 data elements from Archer et al's review in the more recent PHR literature and separated them into more than one data element.

Patient data elements reported in the literature are available in Multimedia Appendix 1.

The comparable data elements identified in both reviews were personal information, problem lists, surgical history (procedures, hospitalizations), medical history (family history), provider information (provider list), allergies, home monitoring data, medical history, psychographics (social history, lifestyle), immunizations, prescription medications, and notes.

The data elements not previously reported in Archer et al and that we identified in this research were (1) genetic data, (2) preferences, (3) PHR settings, (4) facility information, (5) personalized search results, (6) visit preparation information, (7) compliance, (8) medical equipment and supplies, (9) self-treatment, (10) treatment plan, (11) outcomes, (12) patient-provider message, (13) incentive programs data, (14) patient health education material, (15) trainings, (16) personalized health advice, (17) environmental information, (18) assessment information, (19) personal health goals, (20) health care cost management, (21) insurance data, and (22) health status.

In our research, we were also able to separate Archer et al's preventive health recommendations into (1) preventive care and (2) prevention adherence. We broke examinations and diagnoses down into (1) vital signs and anthropometric data, (2) physiological information, and (3) diagnosis. We further distinguished laboratory tests and appointments as (1) results, (2) imaging, and (3) appointments.

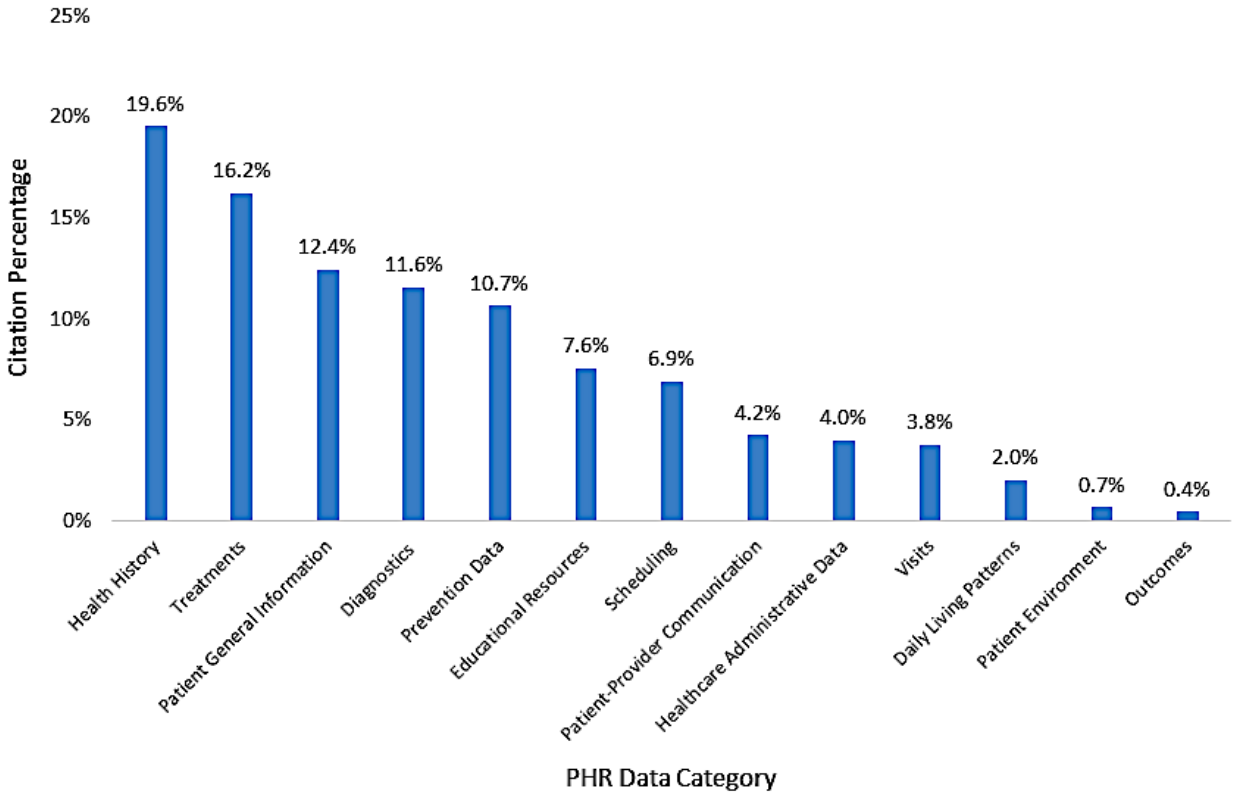
Functional Taxonomy and Chronological Analysis Results

We grouped PHR data elements by source, format, and time of first mention. Analysis of data elements mentioned in the literature allowed for description of information available for analytics use. This analysis also revealed the capabilities available to patients through PHR systems.

Patient Health Record Systems Data—Scope

The bar graph in Figure 2 displays the frequency of data elements described in the articles we reviewed. To obtain the percentages, we divided the total number of citations for each of the major data categories by the total number of citations for all major data categories combined. Figure 2 shows a wide range in the frequency of data categories described in the literature, with health history being the most frequently occurring data category, accounting for 88 out of 450 (19.6%) total citations, and outcomes being the least frequently mentioned, accounting for 2 out of 450 (0.4%) citations.

Figure 2. Patient health record (PHR) data category by citation percentage.



The 4 most frequently occurring data elements (health history, treatments, patient general information, and diagnostics) accounted for 269 out of 450 (59.8%) total citations and were typically added to PHRs through extraction from the patients' EHR.

In addition to data elements extracted from the EHR, a significant amount of data, such as information about medication adherence and self-care, is entered by patients. However, we found PHR-entered data less frequently in our review, representing about 27% of the 450 total citations: 34 (7.6%) citations related to educational resources, 31 (6.9%) citations related to scheduling, 19 (4.2%) citations related to communication, 17 (3.8%) citations related to visits, 9 (2.0%) citations related to daily living patterns, 3 (0.7%) citations related to patient environment, and 2 (0.4%) citations related to outcomes. These likely reflect new functionalities provided to patients through their PHR during our review period. Administrative data accounted for 18 (4.0%) of total citations and consisted of information on health care cost management and insurance data. Health care cost management included information on admissions and discharges and on health spending. Insurance information, on the other hand, provided patients with information such as insurance claims, benefits, copays, and reimbursement.

Data available in the PHR were generated by a multitude of devices, and were entered by different parties (ie, patients and providers) through various platforms (Table 1). We found that data elements related to the patient-provider encounter, such as patient general information, diagnostics, psychosocial status, treatments, visits, and outcomes data, were generally extracted from the EHR. More recent data elements were entered through patient portals (such as educational resources and patient environment data), or transmitted by sensors and tracking devices (such as daily living patterns).

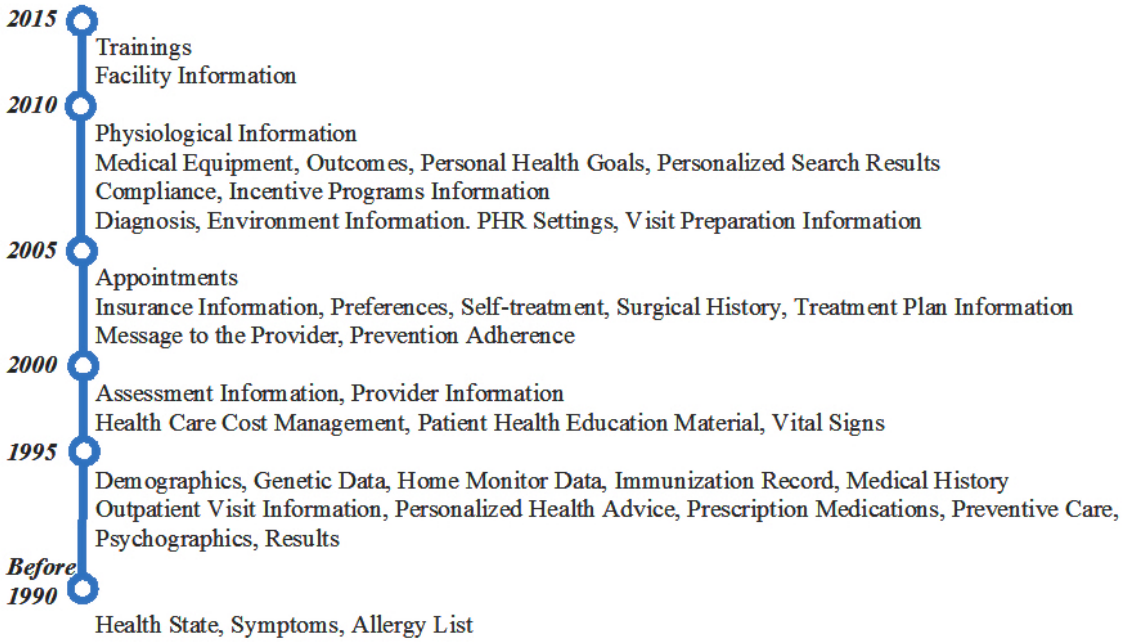
The variety of PHR platforms led to the generation of different data formats (Table 1). Newly generated patient data were not limited to plain text and numbers in structured tables. Electronic messages, for example, were composed of text and metadata describing the time of transmission and the identity of sending and receiving parties. Templated documents and forms were used for standard reports such as legal documents, care plans, and insurance reports [46]. Images, also prevalent in PHRs today, were used by patients and providers to capture, store, and transmit health data, such as radiology results (2-dimensional x-rays, 3-dimensional computed tomography scans, positron emission tomography scans, magnetic resonance imaging scans, 4-dimensional beating heart) [84], signs and symptoms (wound images) [91], camera uploads [31], health trends (growth charts) [46], mood graphs [37], blood sugar graphs[99], laboratory flow sheets [31], and legal documentation (power of attorney for children and adolescents) [22]. Audio and video were used to capture phone call content [46] and record visits [46]. Newer data formats generated by patient tools and mobile apps included Google Maps for facility information and Google Calendar entries associated with appointment scheduling [31].

Table 1. Patient health record data: common formats and sources.

Main data source and data category	Main data type						
	Text	Number	Image	Video	Voice	Time series	GIS[a] or map
Electronic health record							
Patient general information	X	X					
Diagnostics	X	X					
Psychosocial status	X	X					
Treatments	X	X					
Visits	X	X					
Outcomes	X	X					
Patient portal or mobile device							
Educational resources	X	X	X	X	X		
Scheduling	X	X					X
Patient environment	X	X					
Patient-provider communication	X	X	X				
Patient-provider communication	X	X	X				
Administrative record							
Administrative data	X	X					
Sensors or tracking devices							
Prevention data	X	X	X	X	X	X	
Daily living patterns	X	X	X	X	X	X	

[a]GIS: geographic information system.

Figure 3. Patient health record (PHR) data elements by year of first mention.



Patient Health Record Systems Data—Evolution Over Time

Next, we analyzed the data elements extracted by the year of first mention (Figure 3). In the early 1990s, PHR data elements mentioned in the literature pertained to researchers' and practitioners' visions of potential future systems. These included general patient data, such as demographics, and medical encounter information, such as visit summary.

After initial uses of PHR systems in the early 2000s, new data elements such as appointments, preferences, and system settings

emerged. More recently, PHR data included reminders (eg, appointment reminders [51,99,101], medication reminders [93,110,114], screening and laboratory work reminders [42,46,110], immunization reminders [29,30,55,57,82,90], preventive care reminders [21,59,60], and health maintenance reminders [82]), in addition to alerts [22,76,77,99], identification of personal health goals [19,24,38-40,43,72,74], and disease prevention [76,77,99,110,115]. Tracking and monitoring data via e-journals [82] and diaries [50] also became available.

Today, PHR data are generated through different tools and devices. Tracking devices, now transmitting time-series PHR data, are used to monitor patients' vital signs, such as blood pressure and glucose level (biomonitoring devices) [74,99], and to detect abnormal events, such as alerts from implantable cardioverter defibrillators [117].

Patient Health Record Systems Functionalities—Scope

PHR data were mainly used to provide added functionalities to patients. The provider search results [20,22,47,49,64], for example, helped patients locate health care providers and health-related services. Similar functionalities enabled patients to obtain health advice from support groups. Other functionalities assisted patients with preparing for medical encounters through visit preparation questionnaires [24,46,66,70-72]. Functionalities such as incentive programs [43,56,66,73,74] empowered patients through self-health monitoring. Finally, a unique PHR data category discovered in our review, environmental information [36,50,56,67], captured community health concerns and environmental domains, which can be linked to functionalities such as assessment of environment-related risk factors and recommendations for preventive care.

Patient Health Record Functionality Evolution Over Time

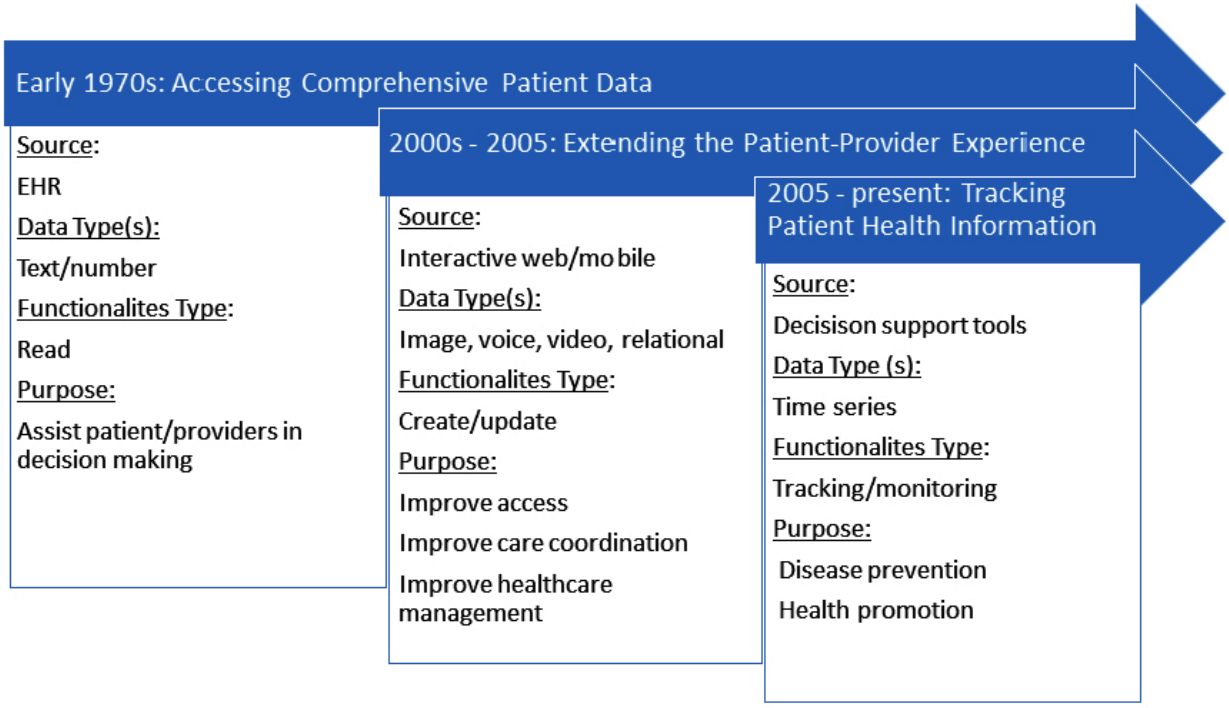
Description of the data extracted revealed which functionalities were available to the patient through the PHR and indicated an interesting evolution of PHR functionalities (Figure 4).

The evolution of PHR data elements over time (Figure 4) illustrates the general inclination in the early stages toward providing the patient with access to health information regarding their medical encounter.

Even though the giving patients access to their own health data was initiated in the 1970s, PHR systems were not widely used until the early 2000s. Because of the infancy of PHR systems, research in this domain has focused on system adoption and how it relates to patient satisfaction. Only limited research is available on how to leverage PHR data to improve health outcomes.

Starting in 2005, data elements reported in the literature indicate a shift toward a more interactive view of the PHR system and the introduction of several new attributes and functionalities. Patient PHR settings, including security and privacy preferences, became more prevalent. The most significant development of this time period of PHR evolution was the interaction and engagement of the patient with the system. Functionalities such as patient-provider secure messaging and appointment scheduling were becoming more common.

Figure 4. Patient health record functionality evolution over time, showing the most common sources, data types, and functionalities found in the review. EHR: electronic health record.



More recently, the PHR system has seen a greater inclusion of patient tracking and monitoring functionalities as daily reported

data from patients and caregivers become more prevalent. Albeit rare, PHR systems also increasingly allow for cost measurement and management.

Discussion

Implications and Future Directions—PHR Data

Overall, the results indicate an increasing focus in the literature on newer types and sources of data, as well as on providing patients with access to their health data. Yet some of these may be progressing so rapidly that important related issues are somewhat neglected. Few studies, for instance, have examined the impact of user interface design on patients' understanding of data and system use. Issues associated with the use of PHRs are mainly related to patients' understanding of the underlying information presented. Problems related to understanding of health data may lead to stress and anxiety [63], which could outweigh the potential benefits of data access. Hence, research is needed in the area of data visualization and representation models specifically targeted for patient use. Examples of such models available in the literature are the what-if analysis, [99] brief intervention [109], and traffic-light feedback system [74]. These methods indicate the risks associated with specific health activities, along with related outcomes and recommended interventions. The traffic-light feedback system, for example, provides patients with an effective visualization tool to track their progress toward attainment of blood pressure goals.

In addition, more research is needed to investigate and improve the quality of patient-entered data. Today, more than 35,000 mobile health apps are available for the iOS and Android operating systems, generating large amounts of data [118]. Data are also increasingly entered through patient forums and portals. While new platforms allow the generation and availability of large data volumes, the wide variety of levels of expertise could lead to reliability and validity issues. Patient-entered data have been shown to be reliable for simple measures such as demographics and symptoms, but less reliable when they pertain to reporting more complicated measures such as laboratory values [5]. One method for improving accuracy could be to provide patients with standardized measures and guidelines for entering their own data, but even that needs to be part of a broader strategy to verify accuracy of data through triangulation from multiple sources.

As the variety of PHR data sources increases, special care is needed for data curation [119] and harmonization [120]. Processes need to be established to produce usable patient-reported data that can be used for research [121]. Standards need to be developed to improve interoperability between different components of the new PHR systems [122]. Data integration methods, such as entity stream mining [123], might be required to cross-reference patient data generated by different tools and devices.

In the coming years, PHR systems will create many data-related challenges, such as quality, heterogeneity, openness, security, scalability, and transparency. Abundant patient data might also trigger information overload. While potentially beneficial for improving health outcomes, streaming patient data can amount to very large volumes, creating new data quality, storage, and analysis issues. All of these challenges open doors for valuable research in health information systems.

The large amounts of data generated by sensors and devices might also require storage and analysis on the cloud [118], potentially increasing storage and analysis costs. Sharing patient data between networks may also create a risk of personal health information disclosure [124], generating additional costs for preserving patient privacy and security. This could also necessitate stronger methods for patient data protection beyond today's practice, which opens up yet another important avenue for health informatics research.

Implications and Future Directions—PHR Functionalities

Overall, PHR data evolution indicates a general trend toward greater patient engagement and health tracking. Moving forward, a continuation of these trends will lead to accumulation of vast amounts of rich data. If patients provide permission, research on PHR data can pave the way for patient-centered care.

The design of patient-centered decision support systems that use a combination of comprehensive individual patient information and aggregate data (collections of patient records) to provide personalized patient recommendations will be a significant area of research.

While past literature has listed patient-provider messaging as an important communication tool for patients and providers, secure message content may potentially provide a valuable patient data source for analysis. Based on their reported intended use, patient secure messages may contain information regarding health-related concerns such as new symptoms and adverse events. Among other possibilities, information retrieved from secure messages could, therefore, be used in research to identify treatment side effects and build patient risk models. However, it is important to keep in mind that terminology used by patients is likely to differ from terminology used by providers. Hence, natural language processing models traditionally used to extract patient information from provider notes may need to be adapted to fit the patient context.

Recently developed and highly effective deep learning algorithms could also be used to extract, search, sort, and analyze information from the tremendous amounts of image, voice, and video data [125] available in the PHR. Other new techniques might be needed to analyze relational data, such as from Google Maps and Google Calendars.

Also, current methods used to store, extract, and analyze EHR data are not adequate for analysis of large volumes of time-series data. Nonrelational databases might be needed to store tracking information. Stream learning algorithms [126] would also need to be applied to extract meaningful information from the terabytes of streaming data analyzed.

As patient-centered decision support systems are being implemented, it is important to ensure the validity of the generated output. Misclassification errors can be dangerous in this domain. Patient systems, which are embedded in mobile devices, need to be evaluated and approved by medical experts.

Data transmitted from different sources can potentially be leveraged by providers to improve patient and population health outcomes. However, accurate measures are still needed to assess and improve the performance of such systems. In addition, these metrics need to account for biases present in patient-generated data. Prior research indicated that PHR systems are mostly used by patients who are typically more sick. Those are patients with comorbidities, such as cancer survivors [127]. Therefore, findings and models generated from analyzing these data might not be generalizable to other patient populations.

The new health care vision in the United States is characterized by automation and collaboration, creating the need for adaptation by all actors in the industry. Empowered patients today have the opportunity to leverage PHR systems data and functionalities. This, however, requires some level of technical expertise for system access and interaction, and medical knowledge in order to understand and interpret the medical information presented. Similarly, medical providers now have to learn and adopt new technologies in order to report medical data and communicate with patients. As a major actor in the health care industry, insurance companies also need to adapt to the new industry environment. Insurance firms today need to assess the value of virtual medical encounters and automated care, and process new types of patient data such as secure messages. Adaptation methods by all health industry players are yet to be assessed and optimized.

Limitations

A limitation of this study is its focus on PHR data reported in the literature. The evolution of PHRs as described in this study might not necessarily reflect the state of the practice. More research is therefore needed to extract and evaluate PHR scope and the functionalities of the various PHR systems available in practice. Also, as mentioned above, this study focused on US studies, thereby limiting the scope of our analysis. Research comparing PHR systems in the United States with those used in other countries would help improve future data uses.

Conclusions

Digital health platforms have changed drastically in recent years. The introduction of distributed PHR systems enabled a shift toward more personalized and increasingly automated health care. The multiuser nature of PHR systems also facilitated patient-to-provider and patient-to-patient information sharing. Yet these changes generated opportunities and challenges at the user, system, and industry levels. Our assessment here of the state of the patient digital infrastructure serves as a valuable foundation for future research. Research implications identified also offer ways to significantly advance health information systems research. Identifying available PHR data also facilitates the development of intelligent health systems. Although primarily aimed at health information systems researchers, implications listed in this study can be further extended to health practitioners, insurance providers, and policy makers.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Patient data elements reported in the literature.

[PDF File (Adobe PDF File), 41KB - jmir_v19i11e388_app1.pdf]

References

1. Shenkin BN, Warner DC. Sounding board. Giving the patient his medical record: a proposal to improve the system. N Engl J Med 1973 Sep 27;289(13):688-692. [doi: 10.1056/NEJM197309272891311] [Medline: 4727972]
2. Hinman E, Holloway J. The patient carried personal health record: a tool to increase patient participation in the treatment process. J Clin Comput 1977;6(4):9.
3. Jackson AN, Kogut S. Use of electronic personal health records to identify patients at risk for aspirin-induced gastrointestinal bleeding. Consult Pharm 2013 May;28(5):313-318. [doi: 10.4140/TCP.n.2013.313] [Medline: 23649680]
4. Mennemeyer ST, Menachemi N, Rahurkar S, Ford EW. Impact of the HITECH Act on physicians' adoption of electronic health records. J Am Med Inform Assoc 2016 Mar;23(2):375-379. [doi: 10.1093/jamia/ocv103] [Medline: 26228764]
5. 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]
6. 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]
7. Gordon MF, Lenderking WR, Duhig A, Chandler J, Lundy JJ, Miller DS, Patient-Reported Outcome Consortium's Cognition Working Group. Development of a patient-reported outcome instrument to assess complex activities of daily living and interpersonal functioning in persons with mild cognitive impairment: The qualitative research phase. Alzheimers Dement 2016 Jan;12(1):75-84. [doi: 10.1016/j.jalz.2015.04.008] [Medline: 26079412]
8. Chung AE, Sandler RS, Long MD, Ahrens S, Burris JL, Martin CF, et al. Harnessing person-generated health data to accelerate patient-centered outcomes research: the Crohn's and Colitis Foundation of America PCORnet Patient Powered

Research Network (CCFA Partners). J Am Med Inform Assoc 2016 May;23(3):485-490 [FREE Full text] [doi: 10.1093/jamia/ocv191] [Medline: 26911821]
9. Bingham CO, Bartlett SJ, Merkel PA, Mielenz TJ, Pilkonis PA, Edmundson L, et al. Using patient-reported outcomes and PROMIS in research and clinical applications: experiences from the PCORI pilot projects. Qual Life Res 2016 Aug;25(8):2109-2116 [FREE Full text] [doi: 10.1007/s11136-016-1246-1] [Medline: 26914103]
10. Segal J. The role of the Internet in doctor performance rating. Pain Physician 2009;12(3):659-664 [FREE Full text] [Medline: 19461833]
11. Heldenbrand S, Martin BC, Gubbins PO, Hadden K, Renna C, Shilling R, et al. Assessment of medication adherence app features, functionality, and health literacy level and the creation of a searchable Web-based adherence app resource for health care professionals and patients. J Am Pharm Assoc (2003) 2016;56(3):293-302. [doi: 10.1016/j.japh.2015.12.014] [Medline: 27067551]
12. 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]
13. Snyder CF, Wu AW, Miller RS, Jensen RE, Bantug ET, Wolff AC. The role of informatics in promoting patient-centered care. Cancer J 2011;17(4):211-218 [FREE Full text] [doi: 10.1097/PPO.0b013e318225ff89] [Medline: 21799327]
14. Markets and Markets. Healthcare analytics/medical analytics market. URL: http://www.marketsandmarkets.com/ResearchInsight/healthcare-data-analytics-market.asp [accessed 2017-11-02] [WebCite Cache ID 6ugNjgKoW]
15. Partnership for the Future of Medicare. Transitioning away from FFS: the plan to start. URL: http://www.futureofmedicare.org/sites/default/files/Policy%20Recommendations%2003.21.13%20.pdf [accessed 2017-11-02] [WebCite Cache ID 6ugNpJJ1V]
16. Marcotte L, Seidman J, Trudel K, Berwick DM, Blumenthal D, Mostashari F, et al. Achieving meaningful use of health information technology: a guide for physicians to the EHR incentive programs. Arch Intern Med 2012 May 14;172(9):731-736. [doi: 10.1001/archinternmed.2012.872] [Medline: 22782203]
17. Wang M, Lau C, Matsen FA, Kim Y. Personal health information management system and its application in referral management. IEEE Trans Inf Technol Biomed 2004 Sep;8(3):287-297. [Medline: 15484434]
18. Williams BT, Imrey H, Williams RG. The lifespan personal health record. Med Decis Making 1991;11(4 Suppl):S74-S76. [Medline: 1770854]
19. Mathis D, O'Reilly K. Point-of-care INR determination, coumadin dosage changes, and use of a historical, self-updating database in a prison. J Correct Health Care 2010 Apr;16(2):139-146. [doi: 10.1177/1078345809356536] [Medline: 20339131]
20. Vawdrey DK, Wilcox LG, Collins SA, Bakken S, Feiner S, Boyer A, et al. A tablet computer application for patients to participate in their hospital care. AMIA Annu Symp Proc 2011;2011:1428-1435 [FREE Full text] [Medline: 22195206]
21. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008;15(1):1-7 [FREE Full text] [doi: 10.1197/jamia.M2562] [Medline: 17947615]
22. Council on Clinical Information Technology. Policy statement--using personal health records to improve the quality of health care for children. Pediatrics 2009 Jul;124(1):403-409. [doi: 10.1542/peds.2009-1005] [Medline: 19564327]
23. Tse CS. A model for developing longitudinal health records. Med Interface 1996 Oct;9(10):68-70, 73. [Medline: 10161513]
24. Cooley Dickinson Hospital. Improving transitions cuts HF readmissions. Hosp Case Manag 2011 Dec;19(12):180-182. [Medline: 22259944]
25. Gerber DE, Laccetti AL, Chen B, Yan J, Cai J, Gates S, et al. Predictors and intensity of online access to electronic medical records among patients with cancer. J Oncol Pract 2014 Sep;10(5):e307-e312 [FREE Full text] [doi: 10.1200/JOP.2013.001347] [Medline: 25006222]
26. Charters K. Challenges of electronic medical record extracts for a personal health record. Stud Health Technol Inform 2009;146:197-201. [Medline: 19592834]
27. Kim E, Wang M, Lau C, Kim Y. Application and evaluation of personal health information management system. Conf Proc IEEE Eng Med Biol Soc 2004;5:3159-3162. [doi: 10.1109/IEMBS.2004.1403891] [Medline: 17270950]
28. Hudson DL, Cohen ME. Uncertainty and complexity in personal health records. Conf Proc IEEE Eng Med Biol Soc 2010;2010:6773-6776. [doi: 10.1109/IEMBS.2010.5625995] [Medline: 21095837]
29. Hogan TP, Wakefield B, Nazi KM, Houston TK, Weaver FM. Promoting access through complementary eHealth technologies: recommendations for VA's Home Telehealth and personal health record programs. J Gen Intern Med 2011 Nov;26 Suppl 2:628-635 [FREE Full text] [doi: 10.1007/s11606-011-1765-y] [Medline: 21989614]
30. Turvey CL, Zulman DM, Nazi KM, Wakefield BJ, Woods SS, Hogan TP, 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]
31. Kharrazi H, Chisholm R, VanNasdale D, Thompson B. Mobile personal health records: an evaluation of features and functionality. Int J Med Inform 2012 Sep;81(9):579-593. [doi: 10.1016/j.ijmedinf.2012.04.007] [Medline: 22809779]
32. Ghossein Y, Gibbs C, Cunningham N, McConathy J, Sung T, Hernandez M, et al. The child's personal health record in New York City: which components are used? Ambul Child Health 1998;4:3-11.

33. Burrington-Brown J, Fishel J, Fox L, Friedman B, Giannangelo K, Jacobs E, et al. Defining the personal health record. AHIMA releases definition, attributes of consumer health record. J AHIMA 2005 Jun;76(6):24-25. [Medline: 15986557]
34. Crouch PB, Rose CD, Johnson M, Janson SL. A pilot study to evaluate the magnitude of association of the use of electronic personal health records with patient activation and empowerment in HIV-infected veterans. PeerJ 2015;3:e852 [FREE Full text] [doi: 10.7717/peerj.852] [Medline: 25802815]
35. Goldzweig C, Towfigh A, Paige N, Orshansky G, Haggstrom D, Beroes J. Systematic Review: Secure Messaging Between Providers and Patients, and Patients' Access to Their Own Medical Record: Evidence on Health Outcomes, Satisfaction, Efficiency and Attitudes. Washington, DC: Department of Veterans Affairs; 2012.
36. Monsen KA, Westra BL, Paitich N, Ekstrom D, Mehle SC, Kaeding M, et al. Developing a personal health record for community-dwelling older adults and clinicians: technology and content. J Gerontol Nurs 2012 Jul;38(7):21-25. [doi: 10.3928/00989134-20120605-03] [Medline: 22715956]
37. Fonda SJ, Kedziora RJ, Vigersky RA, Bursell S. Combining iGoogle and personal health records to create a prototype personal health application for diabetes self-management. Telemed J E Health 2010 May;16(4):480-489. [doi: 10.1089/tmj.2009.0122] [Medline: 20455776]
38. Botts NE, Horan TA, Thoms BP. HealthATM: personal health cyberinfrastructure for underserved populations. Am J Prev Med 2011 May;40(5 Suppl 2):S115-S122. [doi: 10.1016/j.amepre.2011.01.016] [Medline: 21521584]
39. Wagner PJ, Dias J, Howard S, Kintziger KW, Hudson MF, Seol Y, et al. Personal health records and hypertension control: a randomized trial. J Am Med Inform Assoc 2012 Aug;19(4):626-634 [FREE Full text] [doi: 10.1136/amiajnl-2011-000349] [Medline: 22234404]
40. Wagner PJ, Howard SM, Bentley DR, Seol Y, Sodomka P. Incorporating patient perspectives into the personal health record: implications for care and caring. Perspect Health Inf Manag 2010 Oct 01;7:1e [FREE Full text] [Medline: 21063546]
41. Backonja U, Kim K, Casper GR, Patton T, Ramly E, Brennan PF. Observations of daily living: putting the “personal” in personal health records. NI 2012 (2012) 2012;2012:6 [FREE Full text] [Medline: 24199037]
42. Gee PM, Paterniti DA, Ward D, Soederberg MLM. e-Patients perceptions of using personal health records for self-management support of chronic illness. Comput Inform Nurs 2015 Jun;33(6):229-237. [doi: 10.1097/CIN.0000000000000151] [Medline: 25899440]
43. Massoudi BL, Olmsted MG, Zhang Y, Carpenter RA, Barlow CE, Huber R. A web-based intervention to support increased physical activity among at-risk adults. J Biomed Inform 2010 Oct;43(5 Suppl):S41-S45 [FREE Full text] [doi: 10.1016/j.jbi.2010.07.012] [Medline: 20696275]
44. Kupchunas WR. Personal health record: new opportunity for patient education. Orthop Nurs 2007;26(3):185-91; quiz 192. [doi: 10.1097/01.NOR.0000276971.86937.c4] [Medline: 17538475]
45. Chung J, Berkowicz DA, Ho B, Jernigan M, Chueh H. Creating a place for caregivers in personal health: the iHealthSpace copilot program and diabetes care. J Diabetes Sci Technol 2011 Jan 01;5(1):39-46 [FREE Full text] [doi: 10.1177/193229681100500106] [Medline: 21303623]
46. Rocha RA, Romeo AN, Norlin C. Core features of a parent-controlled pediatric medical home record. Stud Health Technol Inform 2007;129(Pt 2):997-1001. [Medline: 17911865]
47. Bouri N, Ravi S. Going mobile: how mobile personal health records can improve health care during emergencies. JMIR Mhealth Uhealth 2014;2(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3017] [Medline: 25098942]
48. Hebda T, Patton C. Application of the relationship-based care model to improve health outcomes via the electronic personal health record. Creat Nurs 2012;18(1):30-33. [Medline: 22482217]
49. Haggstrom DA, Saleem JJ, Russ AL, Jones J, Russell SA, Chumbler NR. Lessons learned from usability testing of the VA's personal health record. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i13-i17 [FREE Full text] [doi: 10.1136/amiajnl-2010-000082] [Medline: 21984604]
50. Lee M, Delaney C, Moorhead S. Building a personal health record from a nursing perspective. Int J Med Inform 2007 Oct;76 Suppl 2:S308-S316. [doi: 10.1016/j.ijmedinf.2007.05.010] [Medline: 17616432]
51. Boland MV, Chang DS, Frazier T, Plyler R, Jefferys JL, Friedman DS. Automated telecommunication-based reminders and adherence with once-daily glaucoma medication dosing: the automated dosing reminder study. JAMA Ophthalmol 2014 Jul;132(7):845-850. [doi: 10.1001/jamaophthalmol.2014.857] [Medline: 24831037]
52. Shimada SL, Brandt CA, Feng H, McInnes DK, Rao SR, Rothendler JA, et al. Personal health record reach in the Veterans Health Administration: a cross-sectional analysis. J Med Internet Res 2014;16(12):e272 [FREE Full text] [doi: 10.2196/jmir.3751] [Medline: 25498515]
53. Lafky DB, Horan TA. Personal health records: consumer attitudes toward privacy and security of their personal health information. Health Informatics J 2011 Mar;17(1):63-71. [doi: 10.1177/1460458211399403] [Medline: 25133771]
54. Greysen SR, Khanna RR, Jacolbia R, Lee HM, Auerbach AD. Tablet computers for hospitalized patients: a pilot study to improve inpatient engagement. J Hosp Med 2014 Jun;9(6):396-399 [FREE Full text] [doi: 10.1002/jhm.2169] [Medline: 24523051]
55. Keith MD, Shimada SL, Rao SR, Quill A, Duggal M, Gifford AL, et al. Personal health record use and its association with antiretroviral adherence: survey and medical record data from 1871 US veterans infected with HIV. AIDS Behav 2013 Nov;17(9):3091-3100. [doi: 10.1007/s10461-012-0399-3] [Medline: 23334359]

56. Fricton J, Davies D. Personal health records to improve health information exchangepatient safety. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in Patient Safety: New Directions and Alternative Approaches. Rockville, MD: Agency for Healthcare Research and Quality; Aug 2008.
57. Nazi KM, Woods SS, Woods SS. MyHealtheVet PHR: a description of users and patient portal use. AMIA Annu Symp Proc 2008:1182. [Medline: 18999142]
58. Tran DT, Zhang X, Stolyar A, Lober WB. Patient-centered design for a personal health record system. AMIA Annu Symp Proc 2005:1140 [FREE Full text] [Medline: 16779426]
59. Mallow JA, Theeke LA, Barnes ER, Whetsel T, Mallow BK. Using mHealth tools to improve rural diabetes care guided by the chronic care model. Online J Rural Nurs Health Care 2014;14(1):43-65 [FREE Full text] [doi: 10.14574/ojrnhc.v14i1.276] [Medline: 26029005]
60. 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]
61. Hagland M. Patient to partner. Will PHRs change the physician-patient relationship? J AHIMA 2005;76(10):38-40. [Medline: 16333942]
62. Felkey B, Fox G. Evaluating personal health record (PHR) options for accountable care organizations (ACOs). Hosp Pharm 2010;45(12):951-952.
63. Kimmel Z, Greenes RA, Liederman E. Personal health records. J Med Pract Manage 2005;21(3):147-152. [Medline: 16471387]
64. Luo G. Open issues in intelligent personal health record--an updated status report for 2012. J Med Syst 2013 Jun;37(3):9943. [doi: 10.1007/s10916-013-9943-6] [Medline: 23584758]
65. Fernandez-Llatas C, Pileggi SF, Ibañez G, Valero Z, Sala P. Cloud computing for context-aware enhanced m-Health services. Methods Mol Biol 2015;1246:147-155. [doi: 10.1007/978-1-4939-1985-7_10] [Medline: 25417085]
66. Buettner K, Fadem SZ. The internet as a tool for the renal community. Adv Chronic Kidney Dis 2008 Jan;15(1):73-82. [doi: 10.1053/j.ackd.2007.10.002] [Medline: 18155112]
67. Killeen J, Chan T, Castillo E, Grisworld W. Integrating environmental data into a personal health record for asthma patients. Ann Emerg Med 2015 Oct;66(4):S101. [doi: 10.1016/j.annemergmed.2015.07.315]
68. Britto MT, Wimberg J. Pediatric personal health records: current trends and key challenges. Pediatrics 2009 Jan;123 Suppl 2:S97-S99. [doi: 10.1542/peds.2008-1755I] [Medline: 19088237]
69. Hagland M. Getting more personal. J AHIMA 2001 Sep;72(8):32-36. [Medline: 15712479]
70. 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]
71. Poon EG, Wald J, Schnipper JL, Grant R, Gandhi TK, Volk LA, et al. Empowering patients to improve the quality of their care: design and implementation of a shared health maintenance module in a US integrated healthcare delivery network. Stud Health Technol Inform 2007;129(Pt 2):1002-1006. [Medline: 17911866]
72. Wald JS, Grant RW, Schnipper JL, Gandhi TK, Poon EG, Businger AC, et al. Survey analysis of patient experience using a practice-linked PHR for type 2 diabetes mellitus. AMIA Annu Symp Proc 2009;2009:678-682 [FREE Full text] [Medline: 20351940]
73. Roett MA, Coleman MT. Practice improvement, part II: update on patient communication technologies. FP Essent 2013 Nov;414:25-31. [Medline: 24261435]
74. Ahern DK, Stinson LJ, Uebelacker LA, Wroblewski JP, McMurray JH, Eaton CB. E-health blood pressure control program. J Med Pract Manage 2012;28(2):91-100. [Medline: 23167022]
75. Dixon BE, Jabour AM, Phillips EO, Marrero DG. An informatics approach to medication adherence assessment and improvement using clinical, billing, and patient-entered data. J Am Med Inform Assoc 2014;21(3):517-521 [FREE Full text] [doi: 10.1136/amiajnl-2013-001959] [Medline: 24076751]
76. Luo G, Tang C, Thomas SB. Intelligent personal health record: experience and open issues. J Med Syst 2012 Aug;36(4):2111-2128. [doi: 10.1007/s10916-011-9674-5] [Medline: 21399913]
77. Demiris G, Thompson H. Smart homes and ambient assisted living applications: from data to knowledge-empowering or overwhelming older adults? Contribution of the IMIA Smart Homes and Ambiant Assisted Living Working Group. Yearb Med Inform 2011;6:51-57. [Medline: 21938325]
78. Tiernan L, Solomon C, Kuehl K. Development of a condition specific personal health record for adolescents and adults with congenital heart disease. J Am Coll Cardiol 2011;57(14):E466.
79. Tang PC, Black W, Young CY. Proposed criteria for reimbursing eVisits: content analysis of secure patient messages in a personal health record system. AMIA Annu Symp Proc 2006:764-768 [FREE Full text] [Medline: 17238444]
80. Hall MJ, Herda MM, Handorf EA, Rybak CC, Keleher CA, Siemon M, et al. Direct-to-patient disclosure of results of mismatch repair screening for Lynch syndrome via electronic personal health record: a feasibility study. Genet Med 2014 Nov;16(11):854-861. [doi: 10.1038/gim.2014.42] [Medline: 24784156]

81. 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]
82. Wright A, Poon EG, Wald J, Feblowitz J, Pang JE, Schnipper JL, et al. Randomized controlled trial of health maintenance reminders provided directly to patients through an electronic PHR. J Gen Intern Med 2012 Jan;27(1):85-92 [FREE Full text] [doi: 10.1007/s11606-011-1859-6] [Medline: 21904945]
83. Nokes KM, Hughes V, Santos R, Bang H. Creating a paper-based personal health record for HIV-infected persons. J Assoc Nurses AIDS Care 2012;23(6):539-547 [FREE Full text] [doi: 10.1016/j.jana.2011.11.004] [Medline: 22512926]
84. Mendelson DS, Bak PRG, Menschik E, Siegel E. Informatics in radiology: image exchange: IHE and the evolution of image sharing. Radiographics 2008;28(7):1817-1833. [doi: 10.1148/rg.287085174] [Medline: 18772272]
85. Mendelson D, Carr C, Chang P, Erickson B, Siegel E, Arenson R. The RSNA/NIBIB image share project: from development to initial experience. Am J Roentgenol 2012 May;198(5 supplement). [doi: 10.2214/ajr.198.5_supplement.ajronline_198_5_supplement_004]
86. Kellick K. The VA anticoagulation management tool. Am J Hematol 2012;87:S150.
87. Gordon J. Educating the patient: challenges and opportunities with current technology. Nurs Clin North Am 2011 Sep;46(3):341-50, vii. [doi: 10.1016/j.cnur.2011.05.005] [Medline: 21791269]
88. Dalan D. Electronic rhinitis outcomes questionnaire (eROQ) & allergic rhinitis and its impact on asthma (ARIA) rhinitis control test (RCT) correlations during a pollen season using an allergy specific personal health record (PHR) on the Internet in the clinic and patient's home. Ann Allergy Asthma Immunol 2009;103(5):A139.
89. Weinert C, Cudney S, Kinion E. Development of My Health Companion to enhance self-care management of chronic health conditions in rural dwellers. Public Health Nurs 2010;27(3):263-269 [FREE Full text] [doi: 10.1111/j.1525-1446.2010.00852.x] [Medline: 20525099]
90. Chumbler NR, Haggstrom D, Saleem JJ. Implementation of health information technology in Veterans Health Administration to support transformational change: telehealth and personal health records. Med Care 2011 Dec;49 Suppl:S36-S42. [doi: 10.1097/MLR.0b013e3181d558f9] [Medline: 20421829]
91. Wilkins EG, Lowery JC, Goldfarb S. Feasibility of virtual wound care: a pilot study. Adv Skin Wound Care 2007 May;20(5):275-6, 278. [doi: 10.1097/01.ASW.0000269315.30639.82] [Medline: 17473564]
92. 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]
93. Krist AH, Peele E, Woolf SH, Rothemich SF, Loomis JF, Longo DR, et al. Designing a patient-centered personal health record to promote preventive care. BMC Med Inform Decis Mak 2011;11:73 [FREE Full text] [doi: 10.1186/1472-6947-11-73] [Medline: 22115059]
94. Jones DA, Shipman JP, Plaut DA, Selden CR. Characteristics of personal health records: findings of the Medical Library Association/National Library of Medicine Joint Electronic Personal Health Record Task Force. J Med Libr Assoc 2010 Jul;98(3):243-249 [FREE Full text] [doi: 10.3163/1536-5050.98.3.013] [Medline: 20648259]
95. Widmer C, Deshazo JP, Bodurtha J, Quillin J, Creswick H. Genetic counselors' current use of personal health records-based family histories in genetic clinics and considerations for their future adoption. J Genet Couns 2013 Jun;22(3):384-392 [FREE Full text] [doi: 10.1007/s10897-012-9557-z] [Medline: 23242928]
96. Østerlund CS, Dosa NP, Arnott SC. Mother knows best: medical record management for patients with spina bifida during the transition from pediatric to adult care. AMIA Annu Symp Proc 2005:580-584 [FREE Full text] [Medline: 16779106]
97. Schnipper J, Volk L, Wald J, Gandhi T, Williams D, Middleton B. A personal health record module improves documentation of family history. J Gen Intern Med 2011;26:S166.
98. Peterson R. My Health Folders. J Am Health Inf Manage Assoc 2006;77(9):97-99.
99. Fonda SJ, Kedziora RJ, Vigersky RA, Bursell S. Evolution of a web-based, prototype personal health application for diabetes self-management. J Biomed Inform 2010 Oct;43(5 Suppl):S17-S21 [FREE Full text] [doi: 10.1016/j.jbi.2010.05.006] [Medline: 20937479]
100. Bosl W, Mandel J, Jonikas M, Ramoni RB, Kohane IS, Mandl KD. Scalable decision support at the point of care: a substitutable electronic health record app for monitoring medication adherence. Interact J Med Res 2013 Jul 22;2(2):e13 [FREE Full text] [doi: 10.2196/ijmr.2480] [Medline: 23876796]
101. Zeng K, Bodenreider O, Nelson SJ. Design and implementation of a personal medication record-MyMedicationList. AMIA Annu Symp Proc 2008 Nov 06:844-848 [FREE Full text] [Medline: 18998962]
102. Hornick TR, Higgins PA, Stollings C, Wetzel L, Barzilai K, Wolpaw D. Initial evaluation of a computer-based medication management tool in a geriatric clinic. Am J Geriatr Pharmacother 2006 Mar;4(1):62-69. [doi: 10.1016/j.amjopharm.2006.03.006] [Medline: 16730622]
103. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]

104. Stock R, Mahoney E, Gauthier D, Center L, Minniti M, Scott J, et al. Developing a community-wide electronic shared medication list. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in Patient Safety: New Directions and Alternative Approaches. Rockville, MD: Agency for Healthcare Research and Quality; Aug 2008.
105. Dullabh PM, Sondheimer NK, Katsh E, Evans MA. How patients can improve the accuracy of their medical records. EGEMS (Wash DC) 2014;2(3):1080 [FREE Full text] [doi: 10.13063/2327-9214.1080] [Medline: 25848614]
106. Derreza A. Pharmacist review of personal electronic health records to detect medication-related problems and verify accuracy of patient self-reported health information. J Am Pharm Assoc 2011;51(2):283.
107. Oldfield E. Electronic personal health records: a tech response to readmission woes. Pharm Times 2014 Apr 15;80(4):a.
108. Kloss L. Greater urgency, sharper focus for PHRs. J Am Health Inf Manage Assoc 2005;76(10):27.
109. Dawson RC, Cuca YP, Kamitani E, Eng S, Zepf R, Draughon J, et al. Using interactive web-based screening, brief intervention and referral to treatment in an urban, safety-net HIV clinic. AIDS Behav 2015 Jun;19 Suppl 2:186-193 [FREE Full text] [doi: 10.1007/s10461-015-1078-y] [Medline: 25963770]
110. Sequist TD, Zaslavsky AM, Colditz GA, Ayanian JZ. Electronic patient messages to promote colorectal cancer screening: a randomized controlled trial. Arch Intern Med 2011 Apr 11;171(7):636-641 [FREE Full text] [doi: 10.1001/archinternmed.2010.467] [Medline: 21149743]
111. Otsuka SH, Tayal NH, Porter K, Embi PJ, Beatty SJ. Improving herpes zoster vaccination rates through use of a clinical pharmacist and a personal health record. Am J Med 2013 Sep;126(9):832.e1-832.e6 [FREE Full text] [doi: 10.1016/j.amjmed.2013.02.018] [Medline: 23830534]
112. Hess R, Clark S, Fischer G, Weimer M, Sullivan S, Roberts M. The impact of an active personal health record in patients with high cardiovascular risk: a pragmatic randomized controlled trial. J Gen Intern Med 2014;29:S227-S227.
113. Luque AE, Corales R, Fowler RJ, DiMarco J, van Keken A, Winters P, et al. Bridging the digital divide in HIV care: a pilot study of an iPod personal health record. J Int Assoc Provid AIDS Care 2013;12(2):117-121 [FREE Full text] [doi: 10.1177/1545109712457712] [Medline: 22965693]
114. Popovich ML, Aramini JJ, Garcia M. Immunizations: the first step in a personal health record to empower patients. Stud Health Technol Inform 2008;137:286-295. [Medline: 18560090]
115. Hess R, Fischer GS, Sullivan SM, Dong X, Weimer M, Zeith C, et al. Patterns of response to patient-centered decision support through a personal health record. Telemed J E Health 2014 Nov;20(11):984-989. [doi: 10.1089/tmj.2013.0332] [Medline: 25243350]
116. Chira P, Nugent L, Miller K, Park T, Donahue S, Soni A, et al. Living Profiles: design of a health media platform for teens with special healthcare needs. J Biomed Inform 2010 Oct;43(5 Suppl):S9-12 [FREE Full text] [doi: 10.1016/j.jbi.2010.05.008] [Medline: 20937487]
117. Sami A, Chen E, Daley C, Doshi R, Heral L, Toscos T. Innovation in cardiac care: direct transmission of remote implantable cardioverter defibrillator data to patients through their electronic personal health records. Circulation 2014;130(suppl 2):A16894.
118. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):28-33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
119. Krumholz HM, Terry SF, Waldstreicher J. Data acquisition, curation, and use for a continuously learning health system. JAMA 2016 Oct 25;316(16):1669-1670. [doi: 10.1001/jama.2016.12537] [Medline: 27668668]
120. Richesson RL, Krischer J. Data standards in clinical research: gaps, overlaps, challenges and future directions. J Am Med Inform Assoc 2007;14(6):687-696 [FREE Full text] [doi: 10.1197/jamia.M2470] [Medline: 17712081]
121. Estabrooks PA, Boyle M, Emmons KM, Glasgow RE, Hesse BW, Kaplan RM, et al. Harmonized patient-reported data elements in the electronic health record: supporting meaningful use by primary care action on health behaviors and key psychosocial factors. J Am Med Inform Assoc 2012;19(4):575-582 [FREE Full text] [doi: 10.1136/amiajnl-2011-000576] [Medline: 22511015]
122. Marcos C, González-Ferrer A, Peleg M, Cavero C. Solving the interoperability challenge of a distributed complex patient guidance system: a data integrator based on HL7's Virtual Medical Record standard. J Am Med Inform Assoc 2015 May;22(3):587-599. [doi: 10.1093/jamia/ocv003] [Medline: 25882034]
123. Krempl G, Spiliopoulou M, Stefanowski J, Žliobaite I, Brzeziński D, Hüllermeier E, et al. Open challenges for data stream mining research. SIGKDD Explor Newsl 2014 Sep 25;16(1):1-10. [doi: 10.1145/2674026.2674028]
124. El Emam K, Neri E, Jonker E, Sokolova M, Peyton L, Neisa A, et al. The inadvertent disclosure of personal health information through peer-to-peer file sharing programs. J Am Med Inform Assoc 2010;17(2):148-158 [FREE Full text] [doi: 10.1136/jamia.2009.000232] [Medline: 20190057]
125. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
126. Gama J, Sebastião R, Pereira RP. On evaluating stream learning algorithms. Mach Learn 2013;90(3):317-346.
127. Williamson R, Meacham L, Cherven B, Hassen-Schilling L, Edwards P, Palgon M, et al. Predictors of successful use of a web-based healthcare document storage and sharing system for pediatric cancer survivors: Cancer SurvivorLink™. J Cancer Surviv 2014 Sep;8(3):355-363. [doi: 10.1007/s11764-014-0346-6] [Medline: 24535124]

Abbreviations

EHR: electronic health record
HITECH: Health Information Technology for Economic and Clinical Health
MeSH: Medical Subject Headings
PHR: patient health record

Edited by G Eysenbach; submitted 19.05.17; peer-reviewed by S Shimada, K Nazi, G Luo, G Strudwick; comments to author 12.07.17; revised version received 01.09.17; accepted 03.10.17; published 15.11.17.

Please cite as:
Bouayad L, Ialynytchev A, Padmanabhan B
Patient Health Record Systems Scope and Functionalities: Literature Review and Future Directions
J Med Internet Res 2017;19(11):e388
URL: http://www.jmir.org/2017/11/e388/
doi:10.2196/jmir.8073
PMID:29141839

Original Paper

Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques

Jim Lumsden[1,2], MSc; Andy Skinner[1,2], PhD.; David Coyle[3], PhD; Natalia Lawrence[4], PhD; Marcus Munafo[1,2], PhD

[1]UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom

[2]The MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom

[3]School of Computer Science, University College Dublin, Dublin, Ireland

[4]School of Psychology, College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom

Corresponding Author:
Jim Lumsden, MSc
UK Centre for Tobacco and Alcohol Studies
School of Experimental Psychology
University of Bristol
12a Priory Rd
Bristol, BS8 1TU
United Kingdom
Phone: 44 117 92 88450
Email: jim.lumsden@bristol.ac.uk

Abstract

Background: The prospect of assessing cognition longitudinally and remotely is attractive to researchers, health practitioners, and pharmaceutical companies alike. However, such repeated testing regimes place a considerable burden on participants, and with cognitive tasks typically being regarded as effortful and unengaging, these studies may experience high levels of participant attrition. One potential solution is to gamify these tasks to make them more engaging: increasing participant willingness to take part and reducing attrition. However, such an approach must balance task validity with the introduction of entertaining gamelike elements.

Objective: This study aims to investigate the effects of gamelike features on participant attrition using a between-subjects, longitudinal Web-based testing study.

Methods: We used three variants of a common cognitive task, the Stop Signal Task (SST), with a single gamelike feature in each: one variant where points were rewarded for performing optimally; another where the task was given a graphical theme; and a third variant, which was a standard SST and served as a control condition. Participants completed four compulsory test sessions over 4 consecutive days before entering a 6-day voluntary testing period where they faced a daily decision to either drop out or continue taking part. Participants were paid for each session they completed.

Results: A total of 482 participants signed up to take part in the study, with 265 completing the requisite four consecutive test sessions. No evidence of an effect of gamification on attrition was observed. A log-rank test showed no evidence of a difference in dropout rates between task variants (χ^2_2=3.0, *P*=.22), and a one-way analysis of variance of the mean number of sessions completed per participant in each variant also showed no evidence of a difference ($F_{2,262}$=1.534, *P*=.21, partial η^2=0.012).

Conclusions: Our findings raise doubts about the ability of gamification to reduce attrition from longitudinal cognitive testing studies.

(J Med Internet Res 2017;19(11):e395) doi:10.2196/jmir.8473

KEYWORDS

behavioral research/methods; games, experimental; computers; cognition; Internet; play and playthings/psychology; boredom; task performance and analysis; executive function; inhibition (psychology)

Introduction

The prospect of assessing cognition remotely and longitudinally is attractive to researchers, health practitioners, and pharmaceutical companies alike. However, traditionally, assessments of cognitive functioning are performed in a laboratory or clinical setting, making multiple testing sessions expensive and a burden to both researchers and participants. At the time of writing, the use of Internet-based platforms for crowdsourcing participants, such as Amazon MTurk [1] and Prolific Academic [2], combined with the growing number of platforms for delivering Web-based cognitive assessments, such as Testable [3] and Gorilla [4], has newly given researchers the ability to gather data on large numbers of people within very short time spans [5-8]. These new technologies have allowed psychological experiments and interventions to be delivered via the Web easily and inexpensively [9-11].

However, one issue for Web-based studies (and particularly longitudinal studies) is that they must compete against the wealth of entertainment and distraction available on the Internet to attract and retain their participants. This is made more difficult by the fact that dropping out of a Web-based study is easier than doing so in the laboratory, that is, a participant needs to only close their browser window [12]. Many authors have reported difficulties sustaining participant numbers for the duration of their Internet-based studies [13,14], and reviews of adherence to intervention trials have documented dropout rates of around 50% [15,16], considerably higher than in laboratory studies where dropout rates are around 13% [17]. The gradual reduction in the number of participants who continue to provide study data over time is known as attrition [16,18]. High levels of attrition may cause studies to suffer from smaller than intended sample sizes, incomplete datasets, wasted participant compensation, and potentially biased results [19-21].

Attrition is often characterized as a "lack of participant engagement" [22,23], but the definition of *engagement* is unclear. One potential definition conceptualizes engagement in a twofold sense [24], both referring to participants' subjective experience of taking part in a study (ie, their enjoyment of the procedure) and participants' behavior when interacting with the study (ie, how often they return to the study website, or how quickly they drop out). Under this definition, attrition is a subcomponent of engagement: an objective behavioral measure that could be assumed to relate to the concept of engagement as a whole.

Gamification has been heralded as a potential mechanism for increasing participant engagement with Web-based studies and interventions [25-27]. The premise is that by adding gamelike features (points, graphics, levels, competition, etc) to an otherwise mundane task, we might be able to create a more enjoyable and compelling experience for the user [28-30]. By utilizing games' ability to engage individuals, it may be possible to make the testing experience less burdensome, thereby reducing attrition. In previous studies, self-report questionnaires of participants' enjoyment have found that gamelike experiments are typically rated as more enjoyable than their nongamelike counterparts [25,31-35]. There are also some examples of gamification increasing *objective* measures of engagement, such as number of optional trials completed [34] or the number of optional testing blocks chosen [36].

Two systematic reviews looked at the effect of gamification on engagement with "online programs" (mostly e-learning) [37] and Web-based mental health interventions [38]. Drawing on the data from 15 studies comparing engagement with gamified programs with nongamified programs, Looyestyn et al [37] found medium to large effects of gamification on objective measures of engagement such as time spent using the program, number of website visits, and volume of contributions. In contrast, Brown et al [38] assessed the impact of gamification on adherence to 61 Internet-based mental health interventions and found that not only was gamification applied fairly lightly (most studies used only one gamelike element) but there was also little evidence for its efficacy. These conflicting findings could be the result of the reviews' different scopes, the lack of studies in Brown's review which specifically assessed the impact of gamification on adherence, or the very minimal gamification found to have been applied in the reviewed mental health interventions.

The reluctance of researchers to liberally apply gamification to precisely designed mental health interventions is understandable; any small change might impact the intervention's efficacy. Within our own field of gamified cognitive assessment, efforts to increase participant engagement and reduce attrition must be implemented carefully to avoid introducing additional cognitive load and affecting the cognitive constructs under test, thus invalidating the task. Although some studies have reported a positive effect of game mechanics on participant performance [39-41], others have found evidence that gamelike tests *do not* improve performance, and may in fact worsen it [31-33,42-44]. For example, Katz et al [42] found that adding a point scoring system to a working memory training task negatively impacted the task's ability to train cognition. These contrasted findings are likely due to the diverse range of cognitive tasks being used and the variety of gamification approaches applied to them, thus highlighting the need for research, which systematically manipulates gamification techniques within a single type of task [25].

We previously conducted a study exploring the impact of two simple game mechanics (points and theme) on the data collected by, and subjective participant ratings of, a response inhibition task [43]. The points variant rewarded participants with points in accordance with their performance on the task, whereas the theme variant utilized a variety of narratively themed stimuli and task backgrounds. A nongame variant was included as a control condition. This was comparable with a clinical version of the task, with some minor graphical changes to ensure suitability for online use. We found that points were rated highest of the three variants on a subjective questionnaire of enjoyment and engagement and did not negatively affect participant performance on the test. However, we found that the narratively themed task was less liked and negatively affected participant performance. We also observed ceiling effects on participant accuracy in all three task variants because of the ease of the response inhibition task we used.

In this study, we aimed to investigate whether simple gamification could reduce participant attrition from a Web-based longitudinal cognitive testing study. Building on our previous study, we used three variants of a response inhibition task, but we switched to using the Stop Signal Task (SST) to increase task difficulty and avoid ceiling effects. We used the same gamelike features (nongame, points, and theme) as in the previous study. We aimed to assess the effect of gamification on attrition using a longitudinal design whereby participants signed up to four compulsory test sessions over 4 consecutive days before entering a 6-day voluntary period where they could continue to take part once per day if they desired. Participants were told that they would receive £4 for completing all compulsory sessions and an additional 50p for each optional session they completed.

We hypothesized that the nongame variant would have the highest attrition rate, losing participants quickly once the fourth session was complete. We expected the points variant initially to maintain high numbers before falling rapidly around day 6 and 7. Finally, we expected the theme variant to lose participants steadily at first before stabilizing to a low attrition rate, eventually retaining a higher number of participants than either the nongame or points variants. For more information on why we predicted these hypotheses, see Multimedia Appendix 1.

Methods

Design and Overview

We used a between-subjects repeated measures experimental design that took place online over 4 to 10 days. The independent variable was SST variant (nongame, points, and theme). The dependent variables of interest were participant attrition, scores on a questionnaire of enjoyment and engagement, two pilot objective measures of engagement, and stop signal reaction times (SSRTs). We preregistered the study on the Open Science Framework [45].

Participants and Procedure

Participants were recruited from the user base of Prolific Academic [2], which handles the process of checking inclusion criteria, displaying study information, and participant reimbursement. We required participants to be older than 18 years and to have English as a first language but had no further criteria. Once registered, participants were directed to the *Mindgames platform* where they entered their prolific ID and received a unique link, which they used to access the study thereafter. They were then randomly assigned to a single task variant for the duration of the study and completed a Web-based consent form before the testing commenced.

Participants were required to complete one 10-min session per day for the first 4 days of the study to receive £4 as compensation for their time. If participants dropped out of the study before completing four sessions and did not contact us with a reason (technical difficulties, etc), then they did not receive any compensation. This was made clear on the information sheet, which participants read before they signed up to the study, and on the study website itself. For the first four sessions, participants were sent daily reminders via the Prolific Academic messaging system. On the fourth day, participants were informed that there would be no more reminders, and that they were free to either drop out or continue to take part in the study each day thereafter for up to 6 days, with each additional session earning them 50p, for a total of between £4 and £7.

The appropriate compensation for the optional sessions was determined by way of a pilot study using the nongame variant only. We randomly allocated participants to one of the three levels of compensation, that is, 50p, £1, or £2 per optional session completed (the base compensation was still £4) and found that the average number of sessions completed per participant was 7.1, 8.4, and 9.4, respectively. Given that we anticipated the nongame variant to be the least motivating of the three variants, that we wanted to avoid ceiling effects, and that we wanted to minimize the motivational influence of the compensation, we opted for a reward of 50p per optional session.

Ethics approval was obtained from the Faculty of Science Research Ethics Committee at the University of Bristol (40361), and the study was conducted according to the revised Declaration of Helsinki [46].

Materials

The Mindgames Platform

Aside from participant recruitment, daily reminders, and reimbursement, all other elements of the study were hosted on a custom website [47]. The website was a single-page Web app written in JavaScript, with a JSON-based Firebase database [48] and PixiJS [49] as the 2D renderer. The site opened to a main menu screen from which the participant could view the number of sessions they had completed and the amount of money they had earned so far (Figures 1-3). Participants had access to a *history* screen, which allowed them to view their previous progress and monitor their results over time. Clicking the start button displayed a series of instruction screens, followed by the SST task and a short questionnaire. The session ended on the history screen, and the main menu's *start* button became inactive until midnight that night. Each session took approximately 10 min to complete. On the first day of taking part, participants also completed a short demographic questionnaire, which collected data on age, sex, ethnicity, level of education, and the number of hours spent playing video games each week.

Figure 1. Menu screen of the nongame task variant.

Figure 2. Menu screen of the points task variant.

Figure 3. Menu screen of the theme task variant.

Stop Signal Task: Nongame Variant

The SST measures response inhibition [50,51], a key feature of executive control [52]. It tests the participant's "action restraint" by presenting a series of stimuli to which the participant must respond as quickly as possible but are occasionally required to withhold a response. These *stop trials* are indicated by a visual warning that presented a brief delay after stimulus presentation. The primary outcome measure of the SST is the SSRT, which is the number of milliseconds of warning a participant needs for them to be able to successfully inhibit their planned response [52].

In this study, we decided to use the SST as opposed to the Go-NoGo task from our previous study [43]. We did this because we found many participants to be performing at ceiling in the Go-NoGo task, which limited our ability to detect differences between the task variants. The SST is more challenging than the Go-NoGo task because it dynamically adjusts the task's difficulty to match the inhibitory control of the user, therefore reducing the likelihood of a participant performing at ceiling.

We based our SST on the widely used CANTAB SST [53,54] albeit with a visual rather than auditory stop signal and some graphical upgrades to make the task more suitable for the Web. Each trial began with a fixation cross that was displayed in the middle of the screen, with two colored zones on the left and right of the fixation cross (Figure 4). After 500 ms, a colored circle appeared over the fixation cross and participants had to respond as rapidly as possible by pressing either the left or right arrow key to indicate which colored zone matched the color of the circle (Multimedia Appendix 2). In 25% of trials, white brackets appeared around the circle after it was shown: when this occurred, the subject had to withhold their response and wait until the next trial began (each trial was displayed for 900 ms). If the participant responded before the stop signal was displayed, then the trial was recorded as failed, but white brackets were not displayed. Between each trial, there was a random intertrial interval ranging from 500 to 1000 ms. The delay between the circle onset and the bracket onset is called the stop signal delay (SSD), and was varied according to a four-staircase tracking algorithm, designed to sample across the *inhibition-probability by SSD* space (see Multimedia Appendix 1) [55,56]. The task consisted of five blocks of 48 trials each, with a 10s break between each block. If the participant minimized the browser window or changed tabs, then the task was paused (because of the default way in which timers in JavaScript operate). However, if the browser window was not in focus but was still visible (eg, on a second monitor), then the task was not paused.

In the nongame variant, the participant's history was presented as a list of previous sessions, with median reaction times and estimated SSRTs (Figure 5). Hovering over a column displayed a brief explanation of the variable (eg, "The reaction time column shows the average time in milliseconds, which it took you to respond to the circles appearing in each session.").

Stop Signal Task: Points Variant

The points variant was similar to the nongame variant but with the addition of a points mechanic and the task being framed as a game. Points are a common feature of gamified tasks [25] and are classed as "1st Step" gamification [57]. In our task, the participant's points score was displayed at the bottom of the screen throughout (Figure 6) (Multimedia Appendix 3). The scoring system was very similar to that used in our previous study [43], which in turn was based on that used by Miranda et al [33]. The scoring system also incorporates the findings of Guitart-Masip et al [58] who found that subjects were much more successful in learning active (go) choices when rewarded for them and passive choices (stop) when punished. On each successful nonstop trial, the participant earned points equal to 0.2 × bonus × (800 - reaction time), and the number of points gained was displayed briefly in the intertrial interval. This bonus was a multiplier (×2, ×3, ×4…), which increased by 1 every 3 trials but decreased by 3 when the participant failed a stop trial. The bonus was not lost on stop trials to which the participant responded before the stop signal was displayed (to all appearances, the trial was not a stop trial). On a successful inhibition to a stop signal, the bonus was not lost, but no points were awarded (as there was no reaction time on which to base the score for that trial). Scores were maintained over blocks but not over sessions. The scoring system was outlined to the participants in the instructions for the task.

The participant's history was presented as a list of median reaction times, SSRTs, and scores from each testing session (Figure 7). Additionally, the participant's highest score was saved as a high score and was displayed in the top right-hand corner throughout every testing session.

Stop Signal Task: Theme Variant

The theme variant was similar to the nongame variant but with the addition of a graphical theme and a sense of progression. The task was framed as a game and featured themed graphics and stimuli, with the yellow and blue stimuli being replaced by images of objects, though still predominantly blue or yellow (Figure 8, Multimedia Appendix 4). The task was presented on a series of different graphical backgrounds (Multimedia Appendix 5) but with some shared elements: a conveyor belt on which objects appeared and two bins to the left and right into which these objects were sorted. The stop signal was explained as an automatic *fault detector*, which scanned objects as they sat on the conveyor.

The participant's history was presented as a map (Figure 9), and previous sessions' summary data were displayed when the user hovered over the corresponding icon. Each level on the map had a unique name and thematic instruction text, with the intention of creating an overarching goal, perceptual curiosity, and fostering a sense of participant progression [59-61].

Figure 4. Screenshot of the nongame stop signal task.

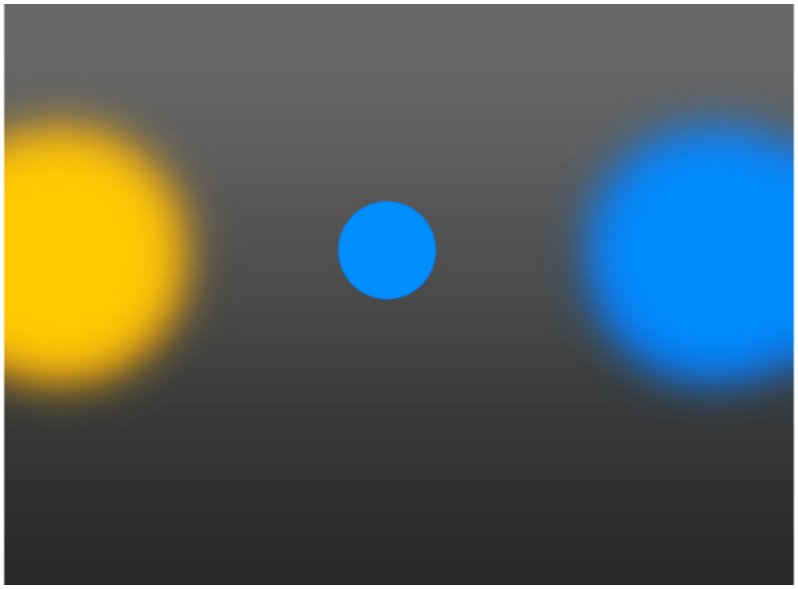


Figure 5. Screenshot of the nongame variant history screen.

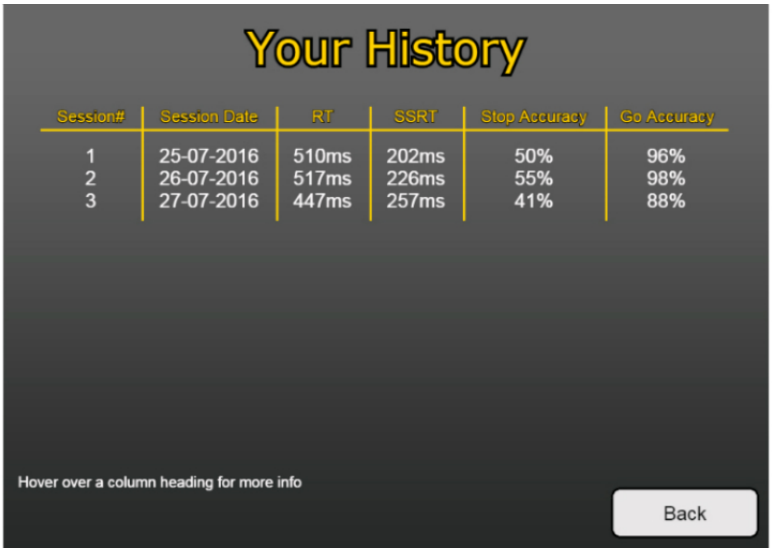


Figure 6. Screenshot of the points variant stop signal task.

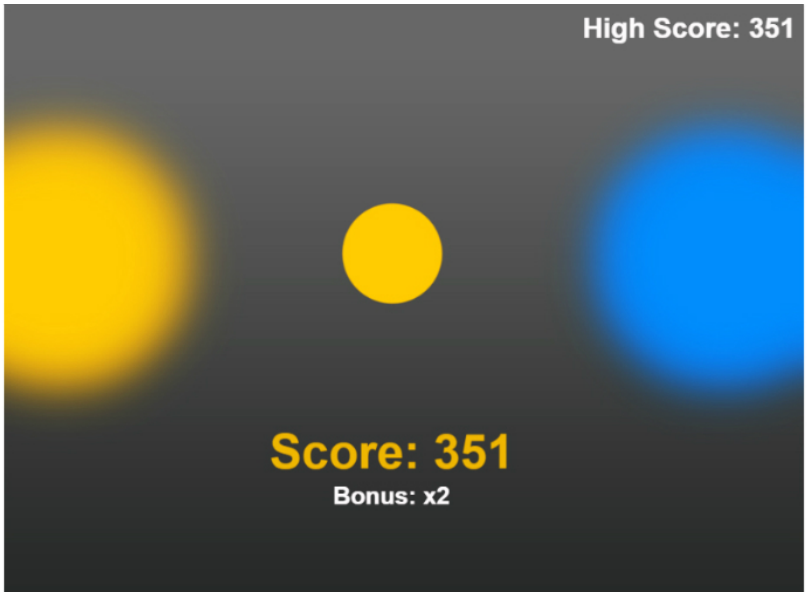


Figure 7. Screenshot of the points variant history screen.

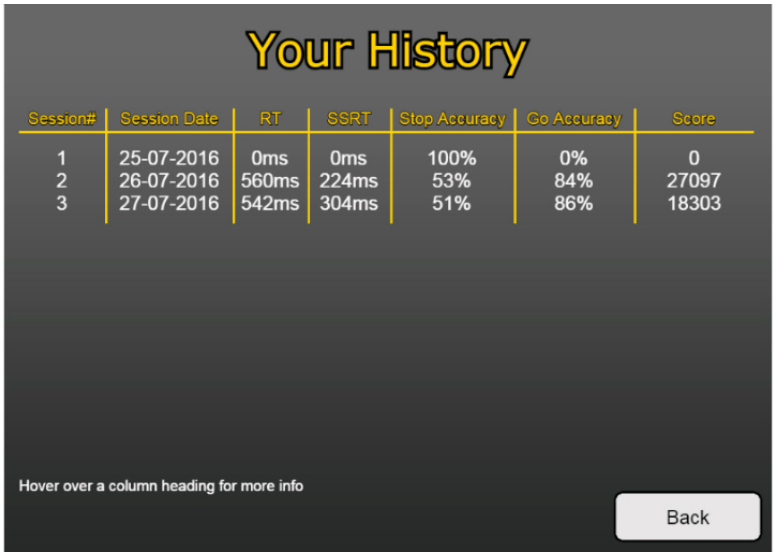


Figure 8. Screenshot of the theme variant stop signal task.

Figure 9. Screenshot of the theme variant history (map) screen.

Enjoyment and Engagement Questionnaire

The enjoyment and engagement questionnaire was designed to collect subjective ratings of the task and was delivered after every session for all three variants. Sessions 1, 4, 7, and 10 delivered the full 10-item questionnaire, whereas the remaining sessions delivered a shorter 5-item questionnaire. These items were answered using a continuous visual analog scale, presented as a horizontal line, 500 pixels long, with a label at either end and no subdivisions. Participants marked a point between these two labels using their mouse.

The following questions were based on those used in previous studies [31,33,43] and were presented in a random order: (1) How enjoyable did you find that?, (2) How frustrating did you find that?, (3) How difficult was it to concentrate for the duration of that?, (4) How well do you think you performed on that?, (5) How mentally stimulating did you find that to be?, (6) How boring did you find that?, (7) How much effort did you put in throughout that?, (8) How repetitive did you find that?, (9) How willing would you be to do that again tomorrow?, and (10) How willing would you be to recommend the study to a friend? Questions 3, 4, 6, 7, and 10 appeared only in the long version of the questionnaire.

Dependent Variable Calculation

Attrition

Attrition was measured in two ways: First, we calculated the mean number of sessions completed per participant (sessions which were started but not finished were excluded from this calculation). Second, we calculated the percentage of participants that completed at least one session, two sessions, etc.

Subjective Measures of Engagement

Subjective engagement with the task was measured by calculating a mean score from the 10-item enjoyment and engagement questionnaire. Questions 2, 3, 6, and 8 were reverse-scored in this calculation. This measure was calculated for each participant's first and fourth sessions, and we also created a *combined score* by taking the mean of the participant's scores from sessions 1 and 4.

Objective Measures of Engagement

We piloted two measures that could potentially serve as objective proxies for engagement: we counted the number of times that participants hid the browser window or moved focus to another window while completing the SST, hypothesizing that unengaged participants would be more likely to briefly visit other websites while testing. We combined the counts of both these events into a single measure: loss-of-focus events. We then created an overall measure of loss-of-focus events for each participant by calculating the mean number of loss-of-focus events from their first four sessions.

We also investigated coefficients of variation, which quantify reaction time intra-individual variability with respect to mean reaction time, as there is some evidence that changes in motivation can be reflected in reaction time variation [62,63]. Coefficients of variation were calculated by dividing the standard deviation (SD) of nonstop trial reaction times by the mean nonstop trial reaction time. Similarly, we created an overall measure of reaction time variation for each participant by calculating the mean coefficient of variation from their first four sessions.

Stop Signal Reaction Times

We calculated SSRTs for each session separately, excluding sessions where the assumptions of the race model did not hold. The race model is a commonly used model of inhibitory control and aims to describe the relationship between stop and go processes [64]. The race model is used to derive the SSRT and so if the assumptions underlying the race model are broken, then the resultant SSRTs are not good representations of the data [50,64]. To that end, we excluded sessions where the median nonstop trial reaction time was longer than the median failed stop-trial reaction time, where SSDs were not positively correlated with their corresponding median failed stop reaction times, and where stop-trial accuracy was not negatively correlated with SSD.

For the sessions that did meet the assumptions of the race model, SSRTs were calculated by modeling an inhibition function, and using it to estimate the SSD at which the participant's probability of inhibiting a stop signal was 50% [56]; we then used this SSD to calculate the SSRT for that session [50,51]. We also created a combined measure of SSRT for each participant by taking the mean SSRT of their first four sessions.

Statistical Analysis

The data that form the basis of our results are available on request from the University of Bristol Research Data Repository [65].

Attrition

Differences in attrition curves were assessed visually using the Kaplan–Meier method to estimate survival functions, a log-rank test, and a one-way analysis of variance (ANOVA) of number of sessions completed.

Subjective Measures of Engagement

We assessed differences in subjective ratings both visually, using bar charts, and using a repeated measures ANOVA of the total score with session number as the time factor and task variant as the between-subjects factor. Where there was evidence of a difference between task variants, we used post hoc *t* tests to investigate further.

Objective Measures of Engagement

We assessed differences in coefficient of variation and website loss-of-focus events between task variants using one-way ANOVAs with data combined across the first four sessions. Where there was evidence of a difference between task variants, we used post hoc *t* tests to investigate further.

Stop Signal Reaction Times

We used boxplots and a one-way ANOVA with task variant as the between-subjects factor to investigate the effects of gamification on SSRT.

Table 1. Interpreting Bayes factors.

Hypothesis 0: the effect size is 0	Strength of evidence	Hypothesis 1: the absolute effect size is between 0 and X[a]
.33≤BF[b]≤1	No support either way	1≤BF≤3
.1≤BF≤.33	Positive	3≤BF≤10
.01≤BF≤.1	Strong	10≤BF≤100
BF<.01	Decisive	BF>100

[a]X: Cauchy prior width.

[b]BF: Bayes factor.

Bayesian Analyses

The three task variants were designed with the aim of minimizing differences in primary task reaction time and nonstop trial accuracy. Therefore, given that frequentist statistics are not ideal for testing equivalences [66,67], we used Bayesian *t* test to assess the evidence for equality of means where frequentist methods failed to find a difference [68,69]. A Bayesian *t* test produces a Bayes Factor (BF), which compares the evidence for two hypotheses. If the evidence favors one hypothesis over the other, then the BF will reflect that, but if the evidence is equal for both hypotheses, then the BF will imply that the data are insensitive [69-71] (Table 1). We used the Bayesian *t* test procedure in JASP [72], with a Cauchy prior width of 0.707. Setting the Cauchy prior width to 0.707 means that in our analysis, one hypothesis is *the effect size is zero* and the other is *the effect size is between −0.707 and 0.707*. Although both hypotheses are centered on an effect size of 0, the former makes a stronger claim than the latter. As such, effect sizes that are not close to 0 are better represented by the latter hypothesis. A prior width of 0.707 was selected for our analysis because it represents the expectation of a medium-large effect, thus weighting the BF against small effects and reducing the likelihood of a false positive.

Sample Size Determination

At the time of study design, to the best of our knowledge, no other studies had investigated the impact of gamification on attrition from a cognitive testing program, and therefore, we had no previous effect size on which to base a sample size determination. Instead, we hypothesized attrition curves (see Multimedia Appendix 1) for each variant and calculated the anticipated effect size (ϕ=0.231) resulting from a Kaplan–Meier method and log-rank test (ie, a chi-square test) on those attrition curves. To detect this difference with alpha=.05 and 95% power, a sample size of 290 was required. We set this to 291 to allow for equal group sizes.

Results

Characteristics of Participants

Participants were recruited in two waves: one starting in October 2016 and another starting in January 2017. In both waves, the intended sample size was met within 3 days of the study being posted on Prolific Academic. A total of 482 participants signed up to take part in the study, with 419 (86.9%) of those completing at least one session. A total of 265 (54.9%) participants completed four sessions over 4 consecutive days as was required by the study criteria (henceforth called *conforming participants*). We excluded 5 participants from the analysis because their reaction times or stimulus categorization accuracy scores were more than 4 interquartile ranges away from the group median. We excluded data from sessions that were started but not completed, and we removed trials from the analysis where participants responded in less than 150 ms.

The analysis below presents data from 260 participants, that is, less than our intended sample size of 291. This was because 32 participants failed to complete the required four sessions in 4 days but instead managed to complete four sessions within 5 days. During the study, we intended on including these *loosely conforming* participants in the analysis, and so stopped recruitment once our intended sample size was achieved. However, for simplicity and adherence to the protocol, we have now decided to present only strictly conforming participant's data below. Analysis of the nonconforming and loosely conforming participants' attrition is presented in Multimedia Appendix 1.

Excluding outliers, 260 conforming participants took part: 91 in the nongame variant, 86 in the points variant, and 83 in the theme. The number of hours spent playing video games was comparable between the groups, and participants typically had a high level of education (Table 2). The most common browser used to complete the experiment was Google Chrome (n=184, 70.8%), with others including Firefox (n=50, 19.2%), Netscape (n=13, 5%), Safari (n=11, 4.2%), Opera (n=1, 0.5%), and Internet Explorer (n=1, 0.5%).

Attrition

Figure 10 shows the attrition of conforming participants, whereas Table 3 shows the mean number of sessions completed per participant in each variant. A log-rank test showed no evidence of a difference between the distributions (χ^2_2=3.0, *P*=.22), and a one-way ANOVA of the number of sessions completed also found no clear evidence of a difference between task variants ($F_{2,262}$=1.534, *P*=.21, partial η^2=0.012). Given the similarity between nongame and points in mean number of sessions completed, we used a Bayesian *t* test to assess their equality and found substantial evidence that they were equal (BF=0.16), but there was no evidence of equality between the theme and the points variant (BF=0.49) or the nongame variant (BF=0.43).

Table 2. Conforming participants' demographic information, shown separately by task variant.

Demographic	Nongame	Points	Theme
Age, mean (SD)	36 (12)	35 (12)	34 (11)
Male, n (%)	43 (47)	49 (57)	42 (51)
Mean video game hours per week (SD)	6 (12)	8 (16)	8 (14)
Median level of education	Bachelor's degree	Bachelor's degree	Bachelor's degree
Mode ethnicity, n (%)	White, 80 (88)	White, 74 (86)	White, 75 (90)

Figure 10. Percentage of conforming participants plotted against the number of sessions they completed, shown separately by task variant.

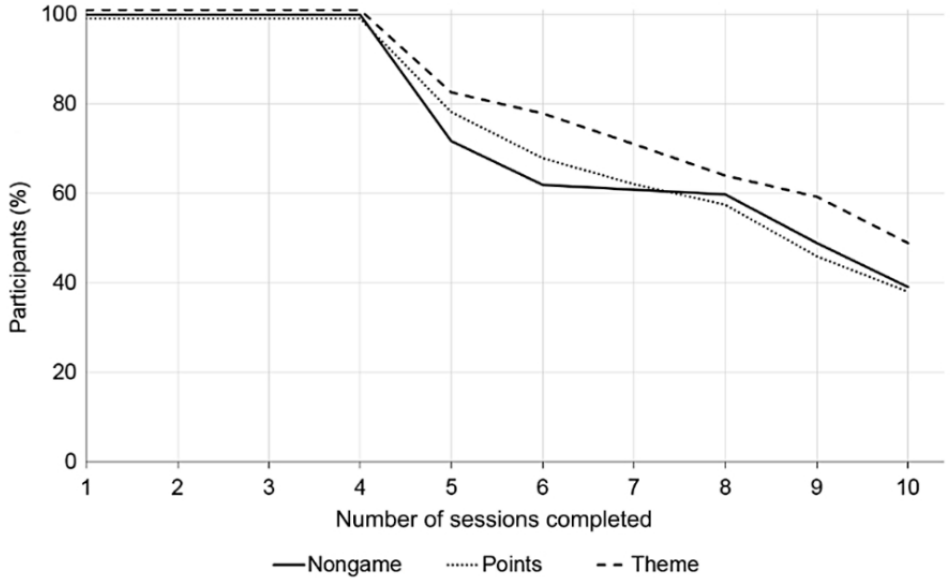


Table 3. Mean number of sessions completed per participant, shown separately by task variant. Conforming participants are those who completed their first four sessions within 4 days as required. All participants includes all who signed up, regardless of their number of sessions completed.

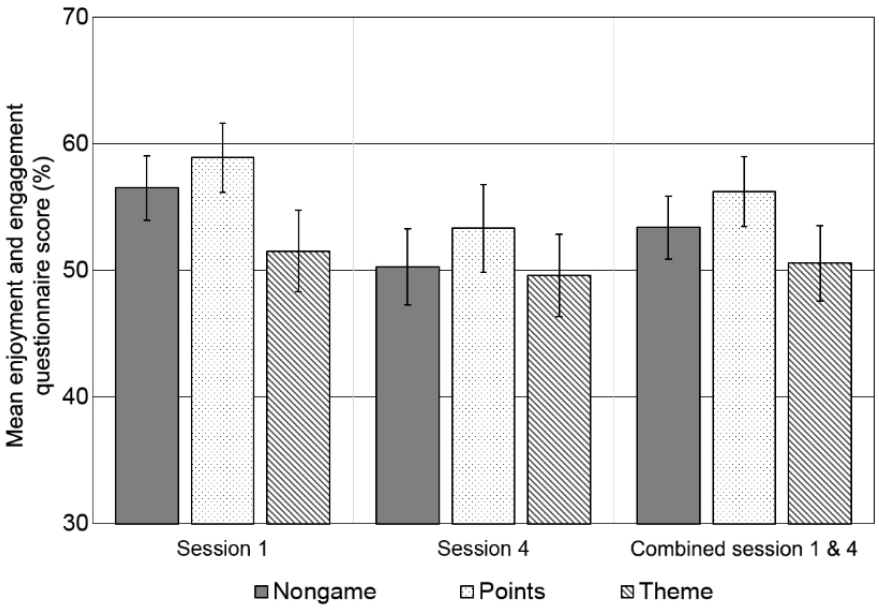
Variant	All participants (95% CI)	Conforming participants (95% CI)
Nongame	4.9 (4.4-5.5)	7.4 (6.8-8.0)
Points	5.1 (4.5-5.6)	7.5 (7.0-8.0)
Theme	5.3 (4.7-5.9)	8.0 (7.5-8.6)

Subjective Measures of Engagement

We used a repeated measures ANOVA of mean score from the enjoyment and engagement questionnaire with session number (1 and 4) as the within-subjects factor and task variant as the between. We used only the two full-length questionnaires completed on the first and the fourth sessions and completed by all participants (for short-form questionnaire results, see Multimedia Appendix 1). We saw evidence for small effects of both task variant ($F_{2,261}$=3.805, P=.02, partial η^2=0.028) and time ($F_{1,261}$=35.693, P<.001, partial η^2=0.120), and weak evidence of an interaction ($F_{2,261}$=3.014, P=.05, partial η^2=0.023). We noted that ratings of all task variants decrease between the first (M=56, 95% CI 54-57) and fourth sessions (M=51, 95% CI 49-53), but it appears that the nongame and points variants were the main drivers of the interaction effect: dropping by 6% (95% CI 4-8) between sessions 1 and 4, whereas ratings of the theme task decreased only by 2% (95% CI −1 to 5). Post hoc t tests on the combined scores showed no evidence for differences between nongame and points, or for nongame and theme (ps>.15) but did show points and theme to be different (mean difference=5.7%, 95% CI 1.6-9.7, t_{171}=2.749, P=.01, d=0.42). Figure 11 shows the mean scores from each task variant at two time points and a combined score taking the averages of both sessions. A breakdown of ratings by individual questions is presented in Multimedia Appendix 1.

As an unplanned exploratory analysis, we were interested in whether a participant's rating on one day predicted their return to the study on the following day. We ran a logistic regression with *returned following day* as the binary dependent variable and the previous day's score on the subjective questionnaire as the predictor variable. However, we saw no evidence that subjective questionnaire scores predicted return the following day (beta=.008; standard error=0.005; $Wald_1$=2.166; P=.14; odds ratio=1.001, 95% CI 0.997-1.019).

Figure 11. Overall scores from the subjective enjoyment and engagement questionnaire. Mean responses of visual analog scale scores from questionnaires delivered on sessions 1 and 4, and the average scores from sessions 1 and 4, shown separately by task variants and time point. Error bars represent 95% CIs.



Objective Measures of Engagement

We analyzed reaction time coefficient of variation and website loss-of-focus events from the four compulsory sessions combined (Table 4). A one-way ANOVA of coefficient of variation showed strong evidence for a medium effect of task variant ($F_{2,260}$=3.131, P=.045, partial η^2=0.024) on participants' reaction time variability, with lower coefficients indicating that there was less variability. Post hoc t tests showed strong evidence of a difference between the points and theme variants (mean difference=1.5%, 95% CI 0.2-2.7; t_{170}=2.349; P=.02; d=0.36), but no clear evidence for other differences were found (ps>.06).

Loss-of-focus events were rare in all task variants, with each participant switching away from the task less than once per session on average. Regardless, we assessed differences in loss-of-focus events between the three task variants using a one-way ANOVA but found no evidence of any difference ($F_{2,260}$=1.137; P=.32; partial η^2=0.008).

Stop Signal Reaction Times

We checked the data from each session against the assumptions of the race model. Of the 1050 sessions assessed, we excluded 161 sessions: 75 from the nongame variant, 37 from points, and 49 from theme. A total of 3 participants failed to meet the assumptions of the race model in all four compulsory sessions, resulting in their exclusion from this analysis. We then analyzed each participant's mean SSRT, with boxplots shown in Figure 12.

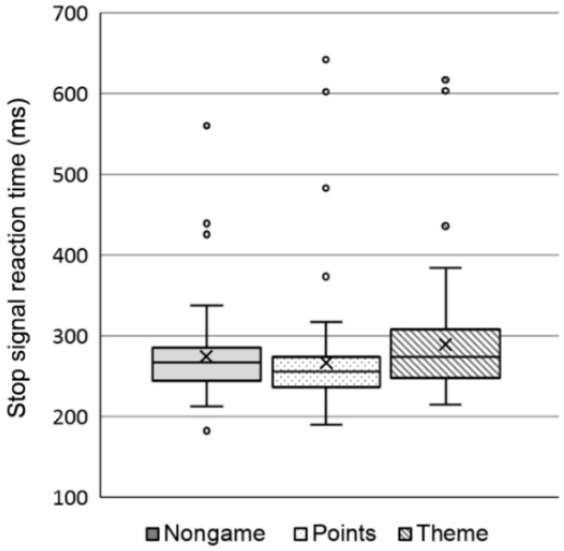
A one-way ANOVA showed weak evidence for a small effect of task variant on SSRT ($F_{2,255}$=2.954; P=.05; partial η^2=0.022) with post hoc t tests showing a difference between the theme variant (M=289; SD=67) and points variant (M=266, SD=66; mean difference=23, 95% CI 5-42; t_{169}=2.386; P=.05; d=0.35). There was no evidence for other differences (ps>.24). Bayesian t tests showed no evidence of equality between the SSRTs of the nongame and theme variants (BF=0.59) but found substantial evidence for equality between the nongame (M=274, SD=55) and the points variants (BF=0.22).

For brevity, not all the analyses planned in the study protocol have been presented. For more detailed methods and analyses, please see Multimedia Appendix 1.

Table 4. Mean objective measures of participant engagement from the first four sessions, shown separately by task variant.

Variant	Coefficient of variation (%) (95% CI)	Loss-of-focus events (95% CI)
Nongame	18.7% (17.9-19.6)	0.85 (0.50-1.19)
Points	19.0% (18.1-19.8)	0.82 (0.43-1.20)
Theme	17.5% (16.7-18.4)	1.21 (0.75-1.67)
Overall	18.4% (17.9-18.9)	0.95 (0.72-1.18)

Figure 12. Boxplots of mean stop signal reaction time. Data combined per participant over the first four sessions and shown separately by task variants.



Discussion

Principal Findings

Contrary to our hypotheses, we saw no clear evidence of an effect of task variant on participant attrition. This was further strengthened when we included data from loosely conforming participants (see Multimedia Appendix 1), which showed strong evidence that the mean number of sessions completed was equal in all task variants. To the best of our knowledge, this is the first empirical study examining the effects of gamification on participant attrition within a cognitive testing context, and our results raise doubts about the efficacy of gamelike tasks for reducing participant dropout.

Despite there being no difference in usage between the variants, we did find an effect of task variant on the subjective ratings of the tasks, with the points variant having the highest *combined sessions* mean, followed by the nongame variant and the theme. One possible explanation for these findings relates to self-determination theory, a popular theory of motivation that centers around the concept of psychological needs and need satisfaction. Self-determination theory posits that human beings have three needs, which are competence, autonomy, and relatedness and that we find activities to be intrinsically motivating if they help us to fulfill these needs [73]. In the case of our gamelike variants, the points variant would seem to address competency needs by providing constant feedback on their performance, which reinforces the player's success [74], but we do not consider the theme variant or the nongame variant to adequately meet any of the three needs. As the points variant was the only variant to address any of these needs, this may explain why it was rated as the most enjoyable in both this study and our previous study [43].

The theme variant was rated as the worst of the three tasks, which was surprising as it maintained the highest percentage of participants until day 10. One potential explanation is that the task was framed as a game and looked like a game but offered no actual gameplay. Moreover, the map screen and changing graphical backgrounds may have hinted at player autonomy and exploration as is typical in other games, but ultimately, the player experience was railroaded. These two factors may have undermined autonomy and violated participant expectations, resulting in a dissatisfying experience [75,76]. Despite this, it is possible that the clear end goal on the map and novelty of changing backgrounds could explain the maintenance of participants in the theme variant while still not being a very satisfying or enjoyable experience.

One additional factor to consider, in the light of self-determination theory, is that paying participants in attrition studies such as this may be counterproductive to measuring true engagement. There is evidence that providing extrinsic rewards for otherwise motivating tasks may undermine participant autonomy, therefore affecting the task's ability to meet our psychological needs [77,78]. In this study, it is not possible to determine whether intrinsic motivation to take part was affected by the incentive of 50p per additional session. This is further complicated by the potentially unrepresentative nature of a Prolific Academic sample: all of whom have voluntarily signed up to take part in science experiments over the Internet but can choose studies based on the amount of monetary compensation awarded in exchange for their data. Given these issues, one potentially informative avenue for future research in this area would be to explore the effects of these same gamification mechanisms on attrition but without providing financial incentives.

Money can be a powerful motivator; for example, Khadjesari et al [79] found that offering a £10 Amazon voucher to each participant in a longitudinal study resulted in a 9% increased response rate. In this study, it may simply be that money was the most important factor for taking part and that the similar attrition rates were driven by the identical incentives.

We also found no evidence that participant ratings of engagement and enjoyment could predict the number of optional sessions they would complete. This, combined with the disconnect between the theme variant ratings and theme variant usage, serves to highlight the split in different types of engagement that has begun to be conceptualized in the literature

[24]. In short, the word engagement has been used in the past to refer to both *engagement as subjective experience* and *engagement as usage*, and this study is further evidence that the two concepts are not as closely related as one might assume. Evidence from the video game literature has found that game enjoyment does not relate strongly to game usage, and that game usage can be driven by many other factors, including boredom, loneliness, and need for escapism [80,81]. This highlights the need for future studies of engagement, which collect both subjective and objective measures.

Our two pilot objective measures of engagement, reaction time variation (coefficients of variation) and loss-of-focus events, were difficult to interpret. We saw no evidence that losses of focus differed between the task variants, and this is likely because such events were rare (less than one loss-of-focus event per session on average). This is a positive finding, as it shows that participants are willing to fully concentrate on Web-based cognitive tasks. With respect to coefficients of variation, the pattern of results is directly in contrast with our subjective measures of engagement: the points variant had the most variable response times but the highest subjective rating, whereas the theme variant had the lowest variability and the lowest rating. This is either evidence contrary to the premise that reaction time variability is related to motivation [62,63] or signals that our subjective ratings are not good measures of motivation. Regardless, further research is necessary to understand whether these objective measures provided are related to the broader concept of engagement.

When assessing cognitive data, we found evidence that SSRTs were equivalent between the points variant and the nongame variant. Although the points variant introduced additional elements to the task, which may have increased cognitive load, it is possible that the highly salient feedback and motivational effect of points served to increase participant performance as has been found in a number of previous studies [41,82-84].

Limitations and Conclusions

First, we consider the fact that we did not achieve our intended sample size, an important limitation of this study. However, we maintain that the results of our supplementary analyses including the loosely conforming participants are quite conclusive and strengthen our finding that there was no effect of gamification on attrition. However, we accept that a balanced group analysis would be preferable. Second, we acknowledge that our sample, recruited from Prolific Academic, with high levels of education, may not be representative of the wider population. Third, we acknowledge that the design of study used is not suitable to validate our gamelike variants as measures of response inhibition, as that would require a within-subjects design to test predictive validity [57,85]. Fourth, the gamelike features we implemented were very lightweight and certainly would not constitute a full game. Indeed, neither of our games were likely enjoyable enough that a participant would consider doing them for their own sake. Though this was necessary to try to reduce the impact of gamification on the cognitive data, it likely reduced any effect of gamification that we might have seen. Fifth, the time course of our study, which took place over days, may not be informative about attrition in studies that take place over longer periods of weeks or months. Sixth, as mentioned previously, there are issues relating to motivation and incentives, as in reality, participants completing cognitive assessments will be presented with requests to complete a study over a fixed period for a fixed fee and not with the option to continue for additional recompense. Finally, both incentives and reminders have been well established as effective methods of increasing engagement, and we used both in this study [86]. Although all three task variants had the same incentives and the same program of reminders (which stopped on day 4), it is possible that these baseline engagement strategies acted as confounders, potentially muddying the effect of gamification on attrition.

In conclusion, the theme variant had negative effects on the cognitive data and showed no clear evidence of reducing attrition. It was also rated as the least enjoyable and was the task switched away from most often. This suggests that themed gamelike tasks, at least those that use graphics alone, are nonoptimal for use in cognitive assessment studies. In contrast, and replicating our previous finding [43], subjective ratings showed the points variant to be well received. We found SSRTs from the points and nongame variants to be equal, showing that points can be an effective way of increasing participant enjoyment of a cognitive task while still collecting valid data.

Despite differences in subjective ratings between the task variants, we saw no effect of gamification on participant attrition over the 6-day optional testing period. Gamification has been promoted as a potential solution to engagement problems in both psychology and digital health care for several years, but we found no effect of gamification on *engagement as usage* in this case. The term gamification may have existed for a decade, but the formalization of gamification's implementation and effectiveness is only just beginning, and there is clearly further work to be conducted to understand how we can translate differences in subjective ratings to differences in usage.

Acknowledgments

The authors gratefully acknowledge the artistic contribution of Melissa Groves who provided graphical resources for the themed task variant. The authors are members of the United Kingdom Centre for Tobacco and Alcohol Studies, a United Kingdom Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UKCRC, is gratefully acknowledged. This work was supported by the Medical Research Council (MC_UU_12013/6 and MC_UU_12013/7) and a PhD studentship to JL funded by the Economic and Social Research Council and Cambridge Cognition Limited. The funders had no role in review design, data extraction and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

JL and MM designed the study. JL programmed the task software and collected the data. JL analyzed the data. AS, NL, and MM contributed to the development of the manuscript.

Conflicts of Interest

DC is a director of Handaxe CIC, a not-for-profit company that develops technology, including computer games, to support mental health interventions for children and adolescents.

Multimedia Appendix 1

A document containing additional details on the methods used in the study and supplementary analyses not core to the message of the paper.

[PDF File (Adobe PDF File), 747KB - jmir_v19i11e395_app1.pdf]

Multimedia Appendix 2

Animated GIF showing several trials of the Non-Game variant of the Stop Signal Task.

[GIF File, 438KB - jmir_v19i11e395_app2.gif]

Multimedia Appendix 3

Animated GIF showing several trials of the Points variant of the Stop Signal Task.

[GIF File, 798KB - jmir_v19i11e395_app3.gif]

Multimedia Appendix 4

Animated GIF showing several trials of the Theme variant of the Stop Signal Task.

[GIF File, 1MB - jmir_v19i11e395_app4.gif]

Multimedia Appendix 5

PDF document showing all possible task backgrounds from the theme variant.

[PDF File (Adobe PDF File), 447KB - jmir_v19i11e395_app5.pdf]

References

1. Mturk. Amazon MTurk URL: https://www.mturk.com/mturk/welcome [accessed 2017-07-11] [WebCite Cache ID 6rsMtCfuJ]
2. Prolific. Bringing people together to power the world's research URL: https://www.prolific.ac/ [accessed 2017-07-11] [WebCite Cache ID 6rsN1syuF]
3. Testable. The web platform for creating, running and sharing behavioral experiments URL: https://www.testable.org/ [accessed 2017-07-11] [WebCite Cache ID 6rsNDoV2R]
4. Gorilla. Behavioural science in the cloud URL: https://gorilla.sc/ [accessed 2017-07-11] [WebCite Cache ID 6rsNEk4Uj]
5. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
6. Woods AT, Velasco C, Levitan CA, Wan X, Spence C. Conducting perception research over the internet: a tutorial review. PeerJ 2015;3:e1058 [FREE Full text] [doi: 10.7717/peerj.1058] [Medline: 26244107]
7. Crump MJ, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One 2013;8(3):e57410 [FREE Full text] [doi: 10.1371/journal.pone.0057410] [Medline: 23516406]
8. 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;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
9. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]
10. Shapiro DN, Chandler J, Mueller PA. Using Mechanical Turk to study clinical populations. Clin Psychol Sci 2013 Jan 31;1(2):213-220. [doi: 10.1177/2167702612469015]
11. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
12. Birnbaum MH. Human research and data collection via the internet. Annu Rev Psychol 2004 Feb;55:803-832. [doi: 10.1146/annurev.psych.55.090902.141601] [Medline: 14744235]

13. Etter JF. Comparing the efficacy of two Internet-based, computer-tailored smoking cessation programs: a randomized trial. J Med Internet Res 2005;7(1):e2 [FREE Full text] [doi: 10.2196/jmir.7.1.e2] [Medline: 15829474]
14. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]
15. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
16. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
17. Lodwick R. Qmro.qmul.ac.: Queen Mary University of London; 2016. Crossover designs: issues in construction, use, and communication URL: http://qmro.qmul.ac.uk/xmlui/handle/123456789/12860 [accessed 2017-09-15] [WebCite Cache ID 6ueOOtap6]
18. Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. Br Med J 2006 Apr 22;332(7547):969-971 [FREE Full text] [doi: 10.1136/bmj.332.7547.969] [Medline: 16627519]
19. Zhou H, Fishbach A. The pitfall of experimenting on the web: how unattended selective attrition leads to surprising (yet false) research conclusions. J Pers Soc Psychol 2016 Oct;111(4):493-504. [doi: 10.1037/pspa0000056] [Medline: 27295328]
20. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
21. 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]
22. Saul JE, Amato MS, Cha S, Graham AL. Engagement and attrition in Internet smoking cessation interventions: insights from a cross-sectional survey of “one-hit-wonders”. Internet Interv 2016 Sep;5:23-29. [doi: 10.1016/j.invent.2016.07.001]
23. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 Steps Australia. J Med Internet Res 2015;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
24. 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;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
25. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
26. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
27. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
28. Deterding S, Sicart M, Nacke L, O'Hara K, Dixon D. Gamification. Using game-design elements in non-gaming contexts. In: CHI '11 Extended Abstracts on Human Factors in Computing Systems. USA: ACM; 2011 Presented at: CHI; May 7-11, 2011; Vancouver, BC, Canada p. A. [doi: 10.1145/1979742.1979575]
29. Rigby S, Ryan RM. Glued to Games: How Video Games Draw Us In and Hold Us Spellbound (New Directions in Media). Santa Barbara, CA: Praeger Publishers Inc; 2010.
30. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]
31. Hawkins GE, Rae B, Nesbitt KV, Brown SD. Gamelike features might not improve data. Behav Res Methods 2013 Jun;45(2):301-318. [doi: 10.3758/s13428-012-0264-3] [Medline: 23055169]
32. McPherson J, Burns NR. Gs invaders: assessing a computer game-like test of processing speed. Behav Res Methods 2007 Nov;39(4):876-883. [Medline: 18183904]
33. Miranda AT, Palmer EM. Intrinsic motivation and attentional capture from gamelike features in a visual search task. Behav Res Methods 2014 Mar;46(1):159-172. [doi: 10.3758/s13428-013-0357-7] [Medline: 23835649]
34. Prins PJ, Dovis S, Ponsioen A, ten BE, van DO. Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? Cyberpsychol Behav Soc Netw 2011 Mar;14(3):115-122. [doi: 10.1089/cyber.2009.0206] [Medline: 20649448]
35. Tong T, Chignell M. Developing a serious game for cognitive assessment: choosing settings and measuring performance. 2014 Apr 26 Presented at: Chinese CHI '14: Proceedings of the Second International Symposium of Chinese CHI; April 26-27, 2014; Toronto, Canada p. 70-79. [doi: 10.1145/2592235.2592246]
36. Dörrenbächer S, Müller PM, Tröger J, Kray J. Dissociable effects of game elements on motivation and cognition in a task-switching training in middle childhood. Front Psychol 2014 Nov 13;5:1275 [FREE Full text] [doi: 10.3389/fpsyg.2014.01275] [Medline: 25431564]

37. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS One 2017;12(3):e0173403 [FREE Full text] [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
38. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
39. Delisle J, Braun C. A context for normalizing impulsiveness at work for adults with attention deficit/hyperactivity disorder (combined type). Arch Clin Neuropsychol 2011 Nov;26(7):602-613. [doi: 10.1093/arclin/acr043] [Medline: 21653627]
40. Dovis S, Van der Oord S, Wiers RW, Prins PJ. Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? The effects of money and computer-gaming. J Abnorm Child Psychol 2012 Jul;40(5):669-681 [FREE Full text] [doi: 10.1007/s10802-011-9601-8] [Medline: 22187093]
41. Ninaus M, Pereira G, Stefitz R, Prada R, Paiva A, Neuper C, et al. Game elements improve performance in a working memory training task. Int J Serious Games 2015 Feb 10;2(1):10. [doi: 10.17083/ijsg.v2i1.60]
42. Katz B, Jaeggi S, Buschkuehl M, Stegman A, Shah P. Differential effect of motivational features on training improvements in school-based cognitive training. Front Hum Neurosci 2014 Apr;8:242 [FREE Full text] [doi: 10.3389/fnhum.2014.00242] [Medline: 24795603]
43. Lumsden J, Skinner A, Woods AT, Lawrence NS, Munafò M. The effects of gamelike features and test location on cognitive test performance and participant enjoyment. PeerJ 2016;4:e2184 [FREE Full text] [doi: 10.7717/peerj.2184] [Medline: 27441120]
44. McPherson J, Burns NR. Assessing the validity of computer-game-like tests of processing speed and working memory. Behav Res Methods 2008 Nov;40(4):969-981. [doi: 10.3758/BRM.40.4.969] [Medline: 19001388]
45. Lumsden J, Munafo M, Lawrence N. Osf. 2016 Oct 11. Effects of game mechanics on participant attrition, data quality and engagement in an online longitudinal cognitive assessment program Internet URL: https://osf.io/ysaqe/ [accessed 2017-11-01] [WebCite Cache ID 6ueQNJIV5]
46. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. J Am Med Assoc 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
47. Lumsden J. Mindgames.firebaseapp. 2016. Mindgames URL: https://mindgames.firebaseapp.com/ [accessed 2017-07-11] [WebCite Cache ID 6rsNag7Fy]
48. Firebase.google. Firebase helps you build better mobile apps and grow your business URL: https://firebase.google.com/ [accessed 2017-11-01] [WebCite Cache ID 6ueSN3dkz]
49. Pixi.js. The HTML5 creation engine URL: http://www.pixijs.com/ [accessed 2017-07-11] [WebCite Cache ID 6rsNHAhV4]
50. Logan GD. On the ability to inhibit thought and action: a users' guide to the stop signal paradigm. In: Inhibitory Processes in Attention, Memory, and Language. San Diego: Academic Press; 1994:189-239.
51. Logan GD, Cowan WB. On the ability to inhibit thought and action: a theory of an act of control. Psychol Rev 1984;91(3):295-327. [doi: 10.1037/0033-295X.91.3.295]
52. Verbruggen F, Logan GD. Response inhibition in the stop-signal paradigm. Trends Cogn Sci 2008 Nov;12(11):418-424 [FREE Full text] [doi: 10.1016/j.tics.2008.07.005] [Medline: 18799345]
53. Cambridgecognition. Cantab research suite URL: http://www.cambridgecognition.com/cantab/ [accessed 2017-11-01] [WebCite Cache ID 6ueSfdK6W]
54. Cambridgecognition. Cantab stop signal task (SST) URL: http://www.cambridgecognition.com/cantab/cognitive-tests/executive-function/stop-signal-task-sst/ [accessed 2017-09-13] [WebCite Cache ID 6tRv9eeZj]
55. Logan GD, Schachar RJ, Tannock R. Impulsivity and inhibitory control. Psychol Sci 2016 May 06;8(1):60-64. [doi: 10.1111/j.1467-9280.1997.tb00545.x]
56. Band GP, van der Molen MW, Logan GD. Horse-race model simulations of the stop-signal procedure. Acta Psychol (Amst) 2003 Feb;112(2):105-142. [doi: 10.1016/S0001-6918(02)00079-3] [Medline: 12521663]
57. Boendermaker WJ, Prins PJ, 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]
58. Guitart-Masip M, Huys QJ, Fuentemilla L, Dayan P, Duzel E, Dolan RJ. Go and no-go learning in reward and punishment: interactions between affect and effect. Neuroimage 2012 Aug 01;62(1):154-166 [FREE Full text] [doi: 10.1016/j.neuroimage.2012.04.024] [Medline: 22548809]
59. Malone T. What makes things fun to learn? Heuristics for designing instructional computer games. USA: ACM; 1980 Presented at: Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems; September 18, 1980; Palo Alto, California p. 162-169.
60. Malone TW. Toward a theory of intrinsically motivating instruction. Cogn Sci 1981 Dec;5(4):333-369. [doi: 10.1016/S0364-0213(81)80017-1]
61. Schell J. The Art of Game Design: A Book of Lenses. San Francisco, California: Morgan Kaufmann; 2008.

62. Andreou P, Neale BM, Chen W, Christiansen H, Gabriels I, Heise A, et al. Reaction time performance in ADHD: improvement under fast-incentive condition and familial effects. Psychol Med 2007 Dec;37(12):1703-1715 [FREE Full text] [doi: 10.1017/S0033291707000815] [Medline: 17537284]
63. Garrett DD, Macdonald SW, Craik FI. Intraindividual reaction time variability is malleable: feedback- and education-related reductions in variability with age. Front Hum Neurosci 2012;6:101 [FREE Full text] [doi: 10.3389/fnhum.2012.00101] [Medline: 22557954]
64. Verbruggen F, Logan GD. Models of response inhibition in the stop-signal and stop-change paradigms. Neurosci Biobehav Rev 2009 May;33(5):647-661 [FREE Full text] [doi: 10.1016/j.neubiorev.2008.08.014] [Medline: 18822313]
65. Lumsden J, Skinner A, Coyle D, Lawrence N, Munafo M. Dataset: No effect of gamification on attrition from a web-based longitudinal cognitive testing study. Bristol: University of Bristol; 2017 Nov 01. [doi: 10.5523/bris.2lh5oz3p93q7p2uyx4w2h70a0v]
66. Berger JO, Sellke T. Testing a Point Null Hypothesis: The Irreconcilability of Values and Evidence. J Am Stat Assoc 1987 Mar 01;82(397):112-122. [doi: 10.1080/01621459.1987.10478397]
67. Blackwelder WC. “Proving the null hypothesis” in clinical trials. Control Clin Trials 1982 Dec;3(4):345-353. [Medline: 7160191]
68. Wetzels R, Raaijmakers JG, Jakab E, Wagenmakers EJ. How to quantify support for and against the null hypothesis: a flexible WinBUGS implementation of a default Bayesian t test. Psychon Bull Rev 2009 Aug;16(4):752-760. [doi: 10.3758/PBR.16.4.752] [Medline: 19648463]
69. Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 2009 Apr;16(2):225-237. [doi: 10.3758/PBR.16.2.225] [Medline: 19293088]
70. Raftery AE. Bayesian model selection in social research. Sociol Methodol 1995;25:111-163. [doi: 10.2307/271063]
71. Jeffreys H. Theory of Probability. Oxford: Oxford University Press; 1998.
72. JASP Team. Jasp-stats. 2017. JASP v0.8.1.2 URL: https://jasp-stats.org/ [accessed 2017-11-01] [WebCite Cache ID 6ueTM2IhH]
73. Ryan RM, Deci EL, editors. Handbook of Self-Determination Research. Rochester: University of Rochester Press; 2002.
74. Sailer M, Hense JU, Mayr SK, Mandl H. How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. Comput Hum Behav 2017 Apr;69:371-380. [doi: 10.1016/j.chb.2016.12.033]
75. Boendermaker WJ, Sanchez Maceiras S, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016 Dec 06;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]
76. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
77. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol Bull 1999 Nov;125(6):627-68; discussion 692. [Medline: 10589297]
78. Deterding S. Situated motivational affordances of game elements: a conceptual model. In: CHI Gamification Workshop 2011. 2011 May 07 Presented at: CHI 2011; 2011; Vancouver, Canada.
79. Khadjesari Z, Murray E, Kalaitzaki E, White IR, McCambridge J, Thompson SG, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. J Med Internet Res 2011;13(1):e26 [FREE Full text] [doi: 10.2196/jmir.1523] [Medline: 21371988]
80. Boyle EA, Connolly TM, Hainey T, Boyle JM. Engagement in digital entertainment games: a systematic review. Comput Hum Behav 2012 May;28(3):771-780. [doi: 10.1016/j.chb.2011.11.020]
81. Lee D, LaRose R. A socio-cognitive model of video game usage. J Broadcast Electron Media 2007;51(4):632-650. [doi: 10.1080/08838150701626511]
82. Attali Y, Arieli-Attali M. Gamification in assessment: do points affect test performance? Comput Educ 2015 Apr;83:57-63. [doi: 10.1016/j.compedu.2014.12.012]
83. Leotti LA, Wager TD. Motivational influences on response inhibition measures. J Exp Psychol Hum Percept Perform 2010 Apr;36(2):430-447 [FREE Full text] [doi: 10.1037/a0016802] [Medline: 20364928]
84. Mekler ED, Brühlmann F, Tuch AN, Opwis K. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Comput Hum Behav 2017 Jun;71(C):525-534. [doi: 10.1016/j.chb.2015.08.048]
85. Kato PM. What do you mean when you say your serious game has been validated? Experimental vs Test Validity. 2013 Apr 25. URL: https://pamkato.com/2013/04/25/what-do-you-mean-when-you-say-your-serious-game-has-been-validated-experimental-vs-test-validty/ [accessed 2016-04-19] [WebCite Cache ID 6gt9POLIu]
86. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

Abbreviations

ANOVA: analysis of variance
BF: Bayes Factor
SD: standard deviation
SSD: stop signal delay
SSRT: stop signal reaction time
SST: Stop Signal Task

Edited by G Eysenbach; submitted 19.07.17; peer-reviewed by W Boendermaker, O Perski; comments to author 03.08.17; revised version received 25.09.17; accepted 09.10.17; published 22.11.17.

Please cite as:
Lumsden J, Skinner A, Coyle D, Lawrence N, Munafo M
Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques
J Med Internet Res 2017;19(11):e395
URL: http://www.jmir.org/2017/11/e395/
doi:10.2196/jmir.8473
PMID:29167090

Original Paper

Performance of a Mobile Phone App-Based Participatory Syndromic Surveillance System for Acute Febrile Illness and Acute Gastroenteritis in Rural Guatemala

Daniel Olson[1,2,3,4], MD; Molly Lamb[2,4], PhD; Maria Renee Lopez[5], MSc; Kathryn Colborn[6,7], PhD; Alejandra Paniagua-Avila[8,9], MD, MPH; Alma Zacarias[8], MD; Ricardo Zambrano-Perilla[10], MS; Sergio Ricardo Rodríguez-Castro[10], MBA; Celia Cordon-Rosales[5], Lic; Edwin Jose Asturias[1,2,3,4], MD

[1]Section of Pediatric Infectious Diseases, University of Colorado School of Medicine, Aurora, CO, United States
[2]Center for Global Health, Colorado School of Public Health, Aurora, CO, United States
[3]Children's Hospital of Colorado, Aurora, CO, United States
[4]Department of Epidemiology, Colorado School of Public Health, Aurora, CO, United States
[5]Centro de Estudios en Salud,, Universidad del Valle de Guatemala, Guatemala City, Guatemala
[6]Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO, United States
[7]Division of Health Care Policy and Research, University of Colorado School of Medicine, Aurora, CO, United States
[8]Fundacion para la Salud Integral de los Guatemaltecos, Center for Human Development, Coatepeque, Guatemala
[9]Center for Public Health Initiatives, Perelman School of Medicine, Philadelphia, PA, United States
[10]Integra IT, Bogota, Colombia

Corresponding Author:
Daniel Olson, MD
University of Colorado School of Medicine
Section of Pediatric Infectious Diseases
13123 East 16th Avenue, Box 055
Aurora, CO, 80045
United States
Phone: 1 7207772838
Email: daniel.olson@ucdenver.edu

Abstract

Background: With their increasing availability in resource-limited settings, mobile phones may provide an important tool for participatory syndromic surveillance, in which users provide symptom data directly into a centralized database.

Objective: We studied the performance of a mobile phone app-based participatory syndromic surveillance system for collecting syndromic data (acute febrile illness and acute gastroenteritis) to detect dengue virus and norovirus on a cohort of children living in a low-resource and rural area of Guatemala.

Methods: Randomized households were provided with a mobile phone and asked to submit weekly reports using a symptom diary app (Vigilant-e). Participants reporting acute febrile illness or acute gastroenteritis answered additional questions using a decision-tree algorithm and were subsequently visited at home by a study nurse who performed a second interview and collected samples for dengue virus if confirmed acute febrile illness and norovirus if acute gastroenteritis. We analyzed risk factors associated with decreased self-reporting of syndromic data using the Vigilant-e app and evaluated strategies to improve self-reporting. We also assessed agreement between self-report and nurse-collected data obtained during home visits.

Results: From April 2015 to June 2016, 469 children in 207 households provided 471 person-years of observation. Mean weekly symptom reporting rate was 78% (range 58%-89%). Households with a poor (<70%) weekly reporting rate using the Vigilant-e app during the first 25 weeks of observation (n=57) had a greater number of children (mean 2.8, SD 1.5 vs mean 2.5, SD 1.3; risk ratio [RR] 1.2, 95% CI 1.1-1.4), were less likely to have used mobile phones for text messaging at study enrollment (61%, 35/57 vs 76.7%, 115/150; RR 0.6, 95% CI 0.4-0.9), and were less likely to access care at the local public clinic (35%, 20/57 vs 67.3%, 101/150; RR 0.4, 95% CI 0.2-0.6). Parents of female enrolled participants were more likely to have low response rate (57.1%, 84/147 vs 43.8%, 141/322; RR 1.4, 95% CI 1.1-1.9). Several external factors (cellular tower collapse, contentious elections) were associated with periods of decreased reporting. Poor response rate (<70%) was associated with lower case reporting

of acute gastroenteritis, norovirus-associated acute gastroenteritis, acute febrile illness, and dengue virus-associated acute febrile illness (P<.001). Parent-reported syndromic data on the Vigilant-e app demonstrated agreement with nurse-collected data for fever (kappa=.57, P<.001), vomiting (kappa=.63, P<.001), and diarrhea (kappa=.61, P<.001), with decreased agreement as the time interval between parental report and nurse home visit increased (<1 day: kappa=.65-.70; ≥2 days: kappa=.08-.29).

Conclusions: In a resource-limited area of rural Guatemala, a mobile phone app-based participatory syndromic surveillance system demonstrated a high reporting rate and good agreement between parental reported data and nurse-reported data during home visits. Several household-level and external factors were associated with decreased syndromic reporting. Poor reporting rate was associated with decreased syndromic and pathogen-specific case ascertainment.

(J Med Internet Res 2017;19(11):e368) doi:10.2196/jmir.8041

KEYWORDS

mobile phone; app; participatory; syndromic surveillance; norovirus; dengue; acute febrile illness; diarrhea; Guatemala

Introduction

Given numerous endemic and emerging infectious diseases, there is a need for improved surveillance capacity in low- and middle-income countries (LMICs). Mobile health (mHealth) systems, which take advantage of the increasing availability of mobile communication technology, offer a potential cost-effective tool to improve surveillance capacity in LMICs by obtaining real-time reporting directly from users.

Health care workers (HCWs) have successfully used mHealth platforms in LMICs for disease surveillance, transmission mapping of pathogens, and decision support [1-5]. However, these systems still rely mostly on passive voluntary reporting at health care facilities, which leads to significant underreporting [6,7]. In addition, the World Health Organization estimates a current shortage of more than 7 million HCWs and this shortage is expected to rise to 12.9 million HCWs by 2035, limiting the capacity of HCWs to report disease [8].

An alternative approach to centralized surveillance is to "task shift" data collection from HCWs and health centers to the health care consumer. Participatory syndromic surveillance systems use mHealth tools for community members to directly self-report symptoms indicative of a particular disease [9,10]. These systems, which have traditionally relied on email- or Internet-based reporting, are now transitioning to short message service (SMS) text messages and, more recently, mobile phone app platforms [9,11]. Participatory syndromic surveillance has been used in developed countries for influenza and syndromic (ie, acute respiratory illness) surveillance in the last decade, but their use in LMICs has been limited, especially in combination with diagnostic testing [9,11-15]. Similar to most LMICs, Guatemala has limited experience with participatory syndromic surveillance and relies almost entirely on passive, centralized surveillance systems to estimate disease burden, including for important emerging pathogens such as dengue virus and norovirus, two of the most common causes of fever and diarrhea, respectively [16-19].

To better understand the utility of participatory syndromic surveillance for emerging disease surveillance in LMICs, we studied the performance and acceptability of a mobile phone app-based participatory syndromic surveillance system in detecting acute febrile illness and acute gastroenteritis among a randomized cohort of children in a low-resource region of Guatemala with limited access to Internet and telecommunication.

Methods

Study Setting and Population

The study was conducted in 25 communities within a 200 km^2 catchment area along the coastal lowlands of southwest Guatemala. The populations living in these communities suffer from high levels of food insecurity, poverty, low access to health care, and high levels of diarrheal and respiratory disease [20]. Households with children aged 6 weeks to 17 years were eligible for study inclusion. All children within a consenting household were offered enrollment, and the household was included if at least one child participated. The household was excluded if the parent was unable to demonstrate proficiency in submitting the weekly mobile phone-based symptom diary after instruction.

Mobile Phone App Development and Use

An encrypted Android mobile phone app (Vigilant-e), developed by Integra IT (Bogota, Colombia), was used. The app allows study participants to directly enter and report symptoms or events using simplified question algorithms and decision-tree logic (see Multimedia Appendix 1 for Vigilant-e app screenshots). Once submitted, encrypted participant data are automatically uploaded using cellular data networks into a secure central database and removed from the mobile phone. Data submitted into the database are available in real time and automated alerts can be generated and sent electronically to study personnel, identifying participants meeting predefined case definitions (acute febrile illness or acute gastroenteritis) or Integrated Management of Childhood Illness (IMCI) warning signs [21].

The Vigilant-e app was configured with the participation of the study investigators, the study nurses, community members, and the Integra IT team. Given low education rates and mobile phone use in the region, the user interface was simplified to a minimum number of questions with simple language and visual aids when possible. Study personnel were trained and the app was field-tested to determine if mobile data coverage was acceptable. Although Internet coverage was variable, the Vigilant-e app was able to store data locally until data coverage was acquired, at which point the data were automatically uploaded into the study database. To avoid unnecessary consumption of available

data from participants, all apps on their mobile phones were blocked except for Vigilant-e and WhatsApp as well as phone and text messaging apps, so participants could communicate with study nurses when needed.

Case Definitions

Case definitions were created prior to study initiation. Acute gastroenteritis was defined as self-reported vomiting or diarrhea for at least 3 days or both for 1 day or more in the preceding week. Norovirus-associated acute gastroenteritis was defined as acute gastroenteritis with concurrent positive norovirus reverse transcription polymerase chain reaction (RT-PCR) testing at the time of sampling. Acute febrile illness was defined as self-reported fever for at least 2 days within the preceding week. Dengue fever was defined as acute febrile illness with positive dengue virus RT-PCR test or anti-DENV IgM serologic test at the time of sampling. If a child reported acute gastroenteritis or acute febrile illness for more than 2 weeks in a row, only the clinical data and sample from the first week were included.

Surveillance System

All households were screened and enrolled using a two-stage cluster sampling strategy adapted from the World Health Organization lot quality assurance method, in which 30 clusters of 7 households were enrolled at random within the study catchment area, as previously described [22-25]. Demographic, geographic information system (GIS), exposure, and clinical data were collected at enrollment, including a survey on the presence and use of mobile phones in the household. All participating households were provided a study-specific mobile phone (Huawei Y330) with the Vigilant-e symptom diary app installed and were asked to complete a symptom report (fever, vomiting, or diarrhea) each week on each participating child. For positive reports of fever, vomiting, or diarrhea, the study nurses phoned the participants to confirm symptoms and ask additional clinical questions. If no weekly report was received from a household in a given week, a study nurse phoned the parent to remind them of the submission requirement. Any participant who met the case definition for acute febrile illness or acute gastroenteritis on their weekly report was visited at their home by a study nurse, with a goal interval time between report and home visit of less than 48 hours. During this visit, study nurses would repeat the same questions asked on the symptom diary app and would obtain a stool or blood sample from participants reporting acute gastroenteritis and acute febrile illness, respectively. Any participant with IMCI danger signs was referred to a medical facility; study physicians supervised all nurses.

Beginning on October 2, 2015, if a participant did not submit the symptom diary in a given week, study nurses were allowed to phone the household and manually enter weekly symptom diary data into the study database instead of relying entirely on the Vigilant-e app. Beginning in early April 2016, all participating households were visited by a study nurse, any malfunctioning phones were replaced, and participants were reminded to submit weekly symptom data using the app. Prospective surveillance using the weekly symptom diary was continued through June 2016, at which point a final closeout visit was performed.

Laboratory Testing

Serum samples from participants with acute febrile illness during prospective follow-up were tested for dengue virus by RT-PCR using the Centers for Disease Control and Prevention's dengue virus assay (DENV-1-4) and IgM anti-DENV IgM enzyme-linked immunosorbent assay (ELISA, InBios Inc, Seattle, WA, USA). Stool specimens were collected using Copan FLOQSwabs (Brescia, Italy) either by rectal swab or on fresh (<2 hours old) stool sample and eluted in eNAT transport solution (Copan, Brescia, Italy) before testing, with both collection techniques previously demonstrating similar molecular viral yield [26,27]. Samples were stored on-site at –20°C and shipped on dry ice to Universidad del Valle de Guatemala (UVG) where diagnostic testing was performed as previously described [25,28].

Statistical Analysis

Demographic variables were compared between the households with 70% or greater response rate and those with less than 70% response rate using a generalized linear model with at binomial response distribution and a log link function. The 70% cutoff was chosen based on response rates observed in previous participatory syndromic surveillance studies in non-LMICs [12,13,29,30]. Agreement of symptoms reported on the Vigilant-e app versus those reported by study nurses at the home visits were calculated using Kendall tau for continuous variables and kappa statistic for categorical variables. SAS v 9.4 (Cary, NC, USA) was used for all data analysis. Mapping of households by response rate was performed using the "sp" spatial package in R and Google satellite images [31].

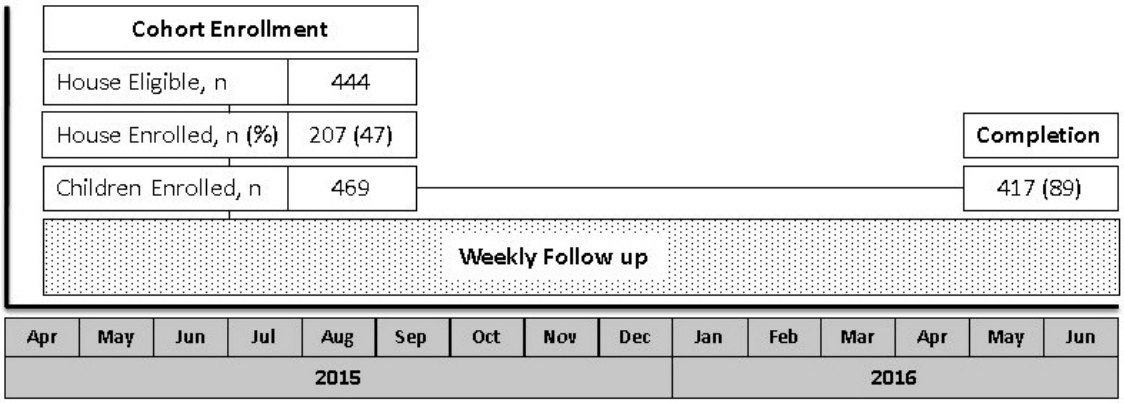
Ethical Oversight

The study was approved by the Colorado Multiple Institutional Review Board, the UVG Institutional Review Board, and the Guatemala Ministry of Health National Ethics Committee. The local Southwest Trifinio Community Advisory Board for Research agreed to the study.

Results

The study enrolled 207 of 444 (46.6%) eligible households from April to September 2015, which included 469 children (Figure 1). The most commonly cited reasons for declining study participation (n=73) included lack of perceived benefit to the child (21/73, 29%), discomfort with rectal specimen collection (16/73, 22%), and not wanting responsibility for the mobile phone (12/73, 16%). Enrolled households were 97.1% (201/207) nonindigenous with a mean of 5.0 (SD 1.8) people living in the household, including mean 2.6 (SD 1.4) children younger than 18 years with a mean age of 7.3 (SD 4.7) years (Table 1). Household density was semirural with a mean of 9.5 (SD 8.6) households per 300 m^2.The primary caregiver (and study mobile phone user in most cases) was literate in 88.4% (183/207) of households; 24.6% (51/207) of fathers and 18.4% (38/207) of mothers had completed secondary education.

Figure 1. Study design and CONSORT diagram of study recruitment, enrollment, and completion. The participatory syndromic surveillance (PSS) cohort enrolled children from April to September 2015, followed by prospective observation with the weekly symptom diary for acute gastroenteritis and acute febrile illness episodes (dotted box) through June 2016.



Cellular phone ownership among households at enrollment varied with 12.1% (25/207) reporting no phone in the household at study initiation and 39.6% (82/207) reporting the use of a mobile phone. Of those households with cellular phones, 72.5% (150/207) used them for text messaging and 37.9% (69/182) for Internet access, of which 19.7% (13/66) accessed the Internet at least daily.

From April 2015 to June 2016, enrolled participants completed 471 person-years of prospective observation with a mean weekly symptom reporting rate of 78% (range 58%-89%) among enrolled children. During the first 25 weeks of observation, study participants with a low (<70%) weekly parental symptom reporting rate using the Vigilant-e app were more likely to be female than participants with a high (≥70%) weekly reporting rate (57% vs 44%, risk ratio [RR] 1.4, 95% CI 1.1-1.9). Households with a low (<70%) weekly symptom reporting rate (n=57) were more likely to have a greater number of children (mean 2.8, SD 1.5 vs mean 2.5, SD 1.3; RR 1.2, 95% CI 1.1-1.4), were less likely to have used mobile phones for text messaging at study enrollment (61% vs 77%, RR 0.6, 95% CI 0.4-0.9), and were less likely to access health care at the local public clinic (35% vs 67%, RR 0.4, 95% CI 0.2-0.6) than households with a ≥70% reporting rate (Table 1). Household cluster density did not differ between households with low versus high reporting rates and there was no significant difference in reporting rates based on geolocation (Figure 2). Households with low weekly reporting rates were significantly less likely to report acute gastroenteritis, norovirus-associated acute gastroenteritis, acute febrile illness, and dengue virus-associated acute febrile illness than households with high weekly reporting (*P*<.01, Table 1). The most commonly cited reasons for not submitting weekly reports were poor data signal, discomfort with using the mobile phone, data consumption, stolen or lost phone, and forgetting to send the report.

Several external factors disrupted the weekly symptom reporting by study participants, including a period of staff turnover, the collapse of a local cellular tower, and contentious primary and secondary national elections (Figure 3). After October 2015, when nurses were allowed to manually enter symptom data directly into the study database for individuals failing to submit symptom data using the Vigilant-e app in a given week, mean overall weekly reporting increased from 73% to 82% over an 8-week period (*P*<.001) (Figure 3). However, during the same timeframe, there was an associated decrease of parents self-reporting symptom data using the Vigilant-e app from 100% to 60% (*P*<.001). After April 2016, when nurses performed the intervention to improve symptom reporting using the Vigilant-e app, the proportion of households self-entering symptom data using the Vigilant-e app increased from 36% preintervention to 54% 4 weeks postintervention (*P*=.046), and the overall mean weekly reporting rate (self-report and nurse report) remained unchanged at 83% (*P*=.93).

When comparing symptom reporting using the Vigilant-e app to nurse-recorded reporting at the home visit (conducted within 48 hours), there was strong agreement between the Vigilant-e app and the home visit for all symptoms except bleeding, which was rarely reported (Table 2). The average time interval between an acute febrile illness or acute gastroenteritis report and a nurse home visit was mean 1.2 (SD 1.7) days; 69.9% (79/113) of participants were visited within 1 day of a report and 86.7% (98/113) were visited within 2 days of a report. The greater the time between the Vigilant-e app symptom report and the nurse home visit, the greater the disagreement in the symptoms (fever, vomiting, and diarrhea) between reporters (Table 3). Overall, at the time of study completion, there was a high satisfaction rate with study participation with 178 (98.8%) of households reporting that participating in the study was beneficial to them personally and 174 (96.6%) reporting that participation benefited the community.

Table 1. Study participant and household characteristics and risk factors associated with low symptom diary app response rate (<70%), April to September 2015.

Characteristics	Overall	Response rate ≥70%	Response rate <70%	RR (CI)[b] (ref ≥70%)	*P*
Child characteristics [a]					
Children enrolled, n	469	322	147		
Age (years), mean (SD)	7.3 (4.7)	7.1 (4.8)	7.5 (4.4)	1.0 (1.0-1.04)	.44
Gender (female), n (%)	225 (47.9)	141 (43.8)	84 (57.1)	1.4 (1.1-1.9)	.008
Child vaccinated (rotavirus), n (%)	250 (53.3)	165 (51.2)	85 (57.8)	1.2 (0.9-1.7)	.13
Child attends school (if age ≥6 years), n (%)	217 (85.1)	143 (84.1)	74 (87.1)	0.8 (0.5-1.4)	.55
Household characteristics					
Households enrolled, n	207	150	57		
Individuals in house, mean (SD)	5.0 (1.8)	4.9 (1.7)	5.2 (2.0)	1.1 (0.9-1.2)	.27
Children enrolled per household, mean (SD)	2.6 (1.4)	2.5 (1.3)	2.8 (1.5)	1.2 (1.1-1.4)	.004
Children aged ≤5 years enrolled per household, mean (SD)	1.0 (0.8)	1.0 (0.8)	1.0 (0.8)	1.0 (0.8-1.4)	.81
Household cluster density, mean (SD)	9.5 (8.6)	9.9 (3.4)	8.1 (8.0)	0.8 (0.6-1.04)	.09
Primary caregiver literacy, n (%)	183 (88.4)	131 (87.3)	52 (91)	1.3 (0.6-3.0)	.52
Father's education ≥secondary, n (%)	51 (24.6)	34 (22.7)	17 (30)	1.3 (0.8-2.1)	.28
Mother's education ≥secondary, n (%)	38 (18.4)	32 (21.3)	6 (11)	0.5 (0.2-1.1)	.10
Health care at public clinic, n (%)	121 (58.5)	101 (67.3)	20 (35)	0.4 (0.2-0.6)	<.001
Duration at current house (years), mean (SD)	8.1 (3.4)	8.04 (3.4)	8.34 (8.8)	1.0 (0.9-1.1)	.63
Cellular phone usage					
Cellular phones per household, mean (SD)	1.4 (1.1)	1.4 (1.0)	1.4 (1.1)	1.0 (0.8-1.2)	.95
Most advanced phone in household, n (%)					
No phone	25 (12.1)	16 (10.7)	9 (16)	Ref	
No mobile phone	99 (47.8)	73 (48.7)	26 (46)	0.7 (0.4-1.4)	.32
Mobile phone	82 (39.6)	60 (40.0)	22 (39)	0.7 (0.4-1.4)	.36
Phones used for text messaging, n (%)[c]	150 (72.5)	115 (76.7)	35 (61)	0.6 (0.4-0.9)	.02
Uses a phone with Internet, n (%)[c]	69 (37.7)	56 (42.1)	13 (27)	0.6 (0.4-1.1)	.10
Internet access frequency[d]**, n (%)**					
≤Weekly	53 (80.3)	45 (81.8)	8 (73)	Ref	
≥Daily	13 (19.7)	10 (17.8)	3 (27)	1.5 (0.5-5.0)	.48
Syndromic reporting (Apr 2015-Jun 2016) [e]					
Acute gastroenteritis, n (%)	100	92 (0.7)	8 (0.3)	0.2 (0.1-0.4)	<.001
Norovirus-associated acute gastroenteritis, n (%)		12 (3.7)	0 (0)	N/C[f]	N/C
Acute febrile illness, n (%)	122	112 (0.9)	10 (0.3)	0.2 (0.1-0.4)	<.001
Dengue-associated acute febrile illness, n (%)		4 (1.2)	0 (0)	N/C	N/C

[a]We were unable to model the random effects of multiple children per household due to relatively low numbers of children per household.

[b]Risk ratios (RR) and 95% confidence intervals were calculated using univariate generalized linear models, with dichotomous response rate in the first 25 weeks of surveillance (≥70% vs <70%) as the outcome of interest.

[c]12% of households are missing these variables.

[d]68% of all households are missing this variable because they said they did not use a phone with Internet access in the previous question.

[e]The response rates reflect the first 25 weeks of surveillance, despite the longer syndromic reporting period (April 2015-June 2016).

[f]N/C: not calculated

Figure 2. Map showing clusters of participants with high (>70%; yellow circle) symptom diary app response rate versus moderate (40%-70%; orange circle) and low (<40%; red circle) response rates, in the Southwest Trifinio Region of Guatemala during the first 25 weeks of surveillance prior to allowing study nurses to manually enter syndromic data (April-October 2015).

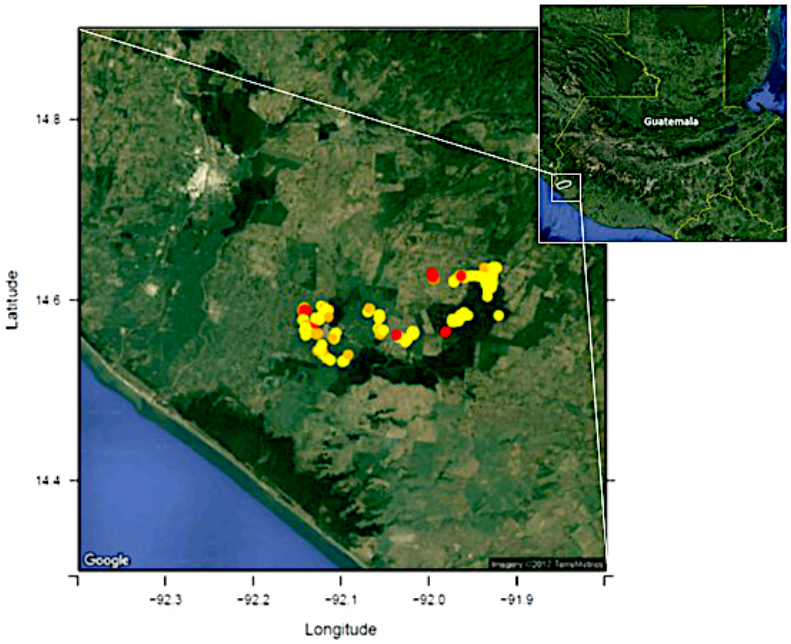


Figure 3. Weekly syndromic reporting rate for acute febrile illness and acute gastroenteritis, April 2015-June 2016. Weekly syndromic reporting rate of participants using the Vigilant-e symptom diary mobile phone app (orange), manually entered data from nurse phone call (green), and combined mobile phone and manual data entry (blue). Several factors were associated with periods of decreased reporting, including a time period of high staff turnover (June-July 2015), a cellular tower collapse (August 2015), and primary and run-off presidential elections (October 2015). On October 2, 2015, study nurses were allowed to manually enter participant data if there was no response. In April 2016, study nurses performed an in-home visit to participating households to repair or replace malfunctioning phones and remind participants to use the Vigilant-e app for reporting when possible.

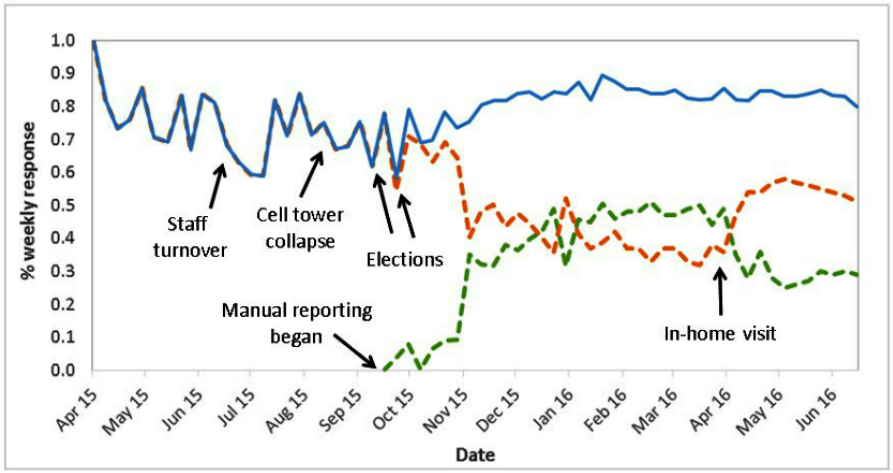


Table 2. Agreement of symptom reporting among study participants between mobile phone symptom diary app and nurse home visit, April 2015 to June 2016.

Symptoms[a]	Symptom diary app (n=113)	Nurse home visit (n=113)	Kappa or Kendall tau[b]	*P* [b]
Fever, n (%)	62 (56.9)	79 (69.9)	.57	<.001
Fever duration (days), mean (SD)	2.9 (1.3)	3.0 (1.8)	.46	<.001
Rash, n (%)	15 (24.2)	16 (20.3)	.59	<.001
Pain, n (%)	38 (61.3)	45 (57.0)	.55	<.001
Nausea, n (%)	29 (46.8)	32 (40.5)	.48	<.001
Bleeding, n (%)	3 (4.8)	1 (1.3)	–.02	.82
Vomiting, n (%)	62 (57.4)	29 (25.7)	.63	<.001
Duration (days), mean (SD)	2.5 (2.0)	1.9 (0.9)	.69	<.001
Maximum emesis/day, mean (SD)	4.5 (2.8)	3.6 (1.7)	.56	.002
Diarrhea, n (%)	33 (30.3)	70 (62.0)	.61	<.001
Diarrhea duration (days), mean (SD)	3.2 (1.8)	3.4 (1.8)	.78	<.001
Maximum stools/day, mean (SD)	4.7 (2.1)	5.1 (2.2)	.29	.006

[a]Participants were asked additional symptom questions if they responded that “yes” their child had fever, diarrhea, or vomiting on the app or the nurse phone call. Nurses also asked the same questions using the same screening technique at the home visit (along with many more detailed questions). Symptoms included any reported symptom, regardless of duration.

[b]Kappa statistic for categorical variables and Kendall tau for continuous variables.

Table 3. Agreement between self-reported symptoms using the Vigilant-e app and study nurse-collected symptoms at home visit.

Days between app report and home visit[a]	n	Fever		Vomiting		Diarrhea	
		Kappa	*P*	Kappa	*P*	Kappa	*P*
<1	79	.70	<.001	.66	<.001	.65	<.001
1	19	.51	.03	.68	.002	.76	.002
≥2	15	.08	.71	.28	.29	.13	.64

[a]As the time interval increased between self-reported symptoms (Vigilant-e app) and nurse-collected symptoms (home visit), agreement between these reporting mechanisms decreased (kappa coefficient). If nurse-collected symptoms occurred within 1 day of self-report, kappa agreement was .65-.70.

Discussion

In a resource-limited region of Guatemala with low literacy rates, we implemented a mobile phone-based participatory syndromic surveillance system with a high weekly response rate and a high rate of agreement between mobile phone parental reporting and nurse home visit reporting. We identified individual and community factors that led to decreased reporting, including female sex of the study participant, a greater number of children in the home, less prior experience with SMS text messaging, and lower utilization of local public health clinics. During our surveillance period, several external factors were associated with decreased reporting, including a cellular tower collapse and national elections, which highlight novel problems in conducting mobile phone-based surveillance. In addition, we demonstrated how contact with study participants, either with phone calls or home visits, may influence self-reporting. Finally, in this region of high norovirus and dengue virus burden [17,25], we found an association between lower rate of weekly self-report and fewer detected episodes meeting syndromic (acute febrile illness and acute gastroenteritis) and pathogen-specific (dengue virus and norovirus) case definitions.

Conducting prospective and accurate infectious disease surveillance in resource-limited settings is difficult for many reasons, including a lack of trained personnel, infrastructure, and diagnostic testing. Prospective surveillance studies have traditionally used weekly home visits or phone calls to collect syndromic data, followed by diagnostic testing in individuals meeting predefined case definitions, although these systems require significant resources and personnel [19,32]. In our study, we task-shifted syndromic data collection directly to community members by training them to use a mobile phone-based symptom diary app. Although study nurses were still required to phone households that did not submit a report in a given week, and we showed that regular contact between study nurses and participants improved reporting, the vast majority of syndromic reporting was entered directly by study participants into an electronic database. The parent-reported data were also accurate when compared to data collected by study nurses, especially when the nurse home visit occurred the same day as the submitted report. Although local syndromic data are unavailable, our estimated acute febrile illness incidence (18.7 per 100 person-years) was similar to that in multiple Latin American countries using a weekly home visit system (26.7 per 100 person-years) [19]. Our estimated acute gastroenteritis incidence

(21.0 per 100 person-years) was lower than estimates reported elsewhere (37-400 episodes per 100 person-years), although these other studies generally included less strict case definitions (1 day of symptoms instead of 3 days), younger populations (<5 years instead of <18 years), and still demonstrated a wide variability in acute gastroenteritis incidence [18,33].

We identified several practical lessons in performing mobile phone-based participatory syndromic surveillance in this resource-limited area. Although mobile phones are increasingly integrating into these communities, the telecommunication infrastructure is susceptible to interference, as demonstrated by decreased reporting following a cellular tower collapse. In addition, other external factors, including a period of high staff turnover that delayed reminder phone calls and widespread protests that impacted both participants and study personnel, were associated with periods of decreased self-reporting. Users were provided with limited data use per month and phones were locked to prevent use of non-study apps, but participants still found ways to circumvent this process. We found that regular communication between study personnel and participants led to improved reporting, but this required more personnel. Allowing study nurses to manually enter participant data into the database, instead of the participants, led to improved reporting overall but was associated with decreased self-reporting, somewhat undermining the participatory syndromic surveillance system. Prior to this intervention, the reporting rate (74%) was consistent with participatory syndromic surveillance studies in non-LMICs [12,13,29,30], and depending on the needs of the surveillance system (eg, vaccine effectiveness, outbreak response), may be sufficient. Further studies are needed to better define the role of participatory syndromic surveillance in these specific settings, and to optimize the interaction between participant and study personnel.

As mobile phones and data networks become increasingly integrated into resource-limited regions of the world, mobile phone-based participatory syndromic surveillance will likely become a more powerful tool to collect population-level syndromic data. Although we provided mobile phones to participants in our study, an important future step will be to allow users to download a symptom diary app onto their own mobile phones and encouraging syndromic self-reporting by providing some small incentive, such as prepaid airtime. This strategy, although still requiring personnel to maintain communication with participants and collect samples when needed, would allow a significant scale-up of the surveillance platform. And although syndromic data would be collected for all individuals, limiting diagnostic testing to a random or higher risk (meeting a more specific case definition) subset of participants would significantly reduce costs. This type of system could also be used for specific populations, such as in screening pregnant women for Zika virus, or during outbreaks of emerging pathogens such as Ebola. Because access to mobile phones is not evenly distributed within a population, it will be important for these types of surveillance programs to find strategies to ensure population-level representativeness.

The study had several strengths and limitations. We chose to perform the study in one of the most resource-limited regions of Guatemala where 70% of the population reports food insecurity and where 60% of households do not own a mobile phone [20]; success in this setting supports the likelihood of replicating this system in other areas with equal or greater resources. Individuals who owned mobile phones at baseline had attained a higher level of education compared to those that did not own mobile phones (P<.001), and they presumably had higher socioeconomic status, so we provided a study mobile phone to all enrolled households to limit bias in our study population. Nevertheless, many eligible households (>50%) declined participation, including 16% that did not want responsibility for the mobile phone, which may have led to a selection bias. We were unable to determine whether the decreased syndromic reporting in the low (<70%) response group was due to actual decreased disease incidence or due to decreased reporting. The representativeness of the system may be biased toward high (≥70%) response households, which is an important consideration for future studies. Sample collection within our prospective syndromic surveillance system decreased over time secondary to parental refusal [25], and although syndromic reporting was maintained at a high level throughout the study, it is possible that households were not reporting symptoms to avoid diagnostic testing (venipuncture or rectal/stool swab). This possibility could have been evaluated by sending study nurses at random to households that reported no symptoms in a given week and comparing their data to the mobile phone self-report. Households did report that the study, including the nurse home visits, benefited their families (98.8%) and the community (96.6%) overall.

In summary, we successfully implemented a mobile phone-based participatory syndromic surveillance system in a resource-limited region of Guatemala and identified several factors that positively or negatively impacted self-reporting. Self-reporting using a symptom diary mobile phone app (Vigilant-e) was accurate when compared to study nurse-collected data during a home visit. Future studies should evaluate mobile phone-based participatory syndromic surveillance for specific high-risk populations and pathogens at other sites, and should scale-up syndromic self-reporting with diagnostic testing performed only within a randomized or select subset of responders.

Acknowledgments

We thank the families who participated in the study and the community leaders who facilitated its implementation. In addition, we are grateful to the following for their significant contributions to this research: CU Trifinio Research Team, including Neudy Carolina Rojop, Carmen Andrea Chacon, Jeniffer Yajaira Cardenas, Edwin Estuardo Hernandez, Edgar Eduardo Barrios, Macaria Genoveva Bail, Ruth Aide Ramirez Angel, Maria Eloin Dhaenes Vivar, Dulce Maria Camas, and Carlos Alvarez Guillen; Universidad del Valle de Guatemala: Mirsa Ariano and Erick Mollinedo; and Integra IT Colombia. This study was supported by

an Investigator-Initiated Sponsored Research Grant from Takeda Pharmaceuticals (IISR-2014-100647). Dr Olson is supported by NIH/NCATS Colorado CTSI Grant Number UL1 TR001082 and the Children's Hospital of Colorado Research Scholar Award. Contents are the authors' sole responsibility and do not necessarily represent official NIH view.

Authors' Contributions

Study design, execution, and interpretation of data was done by DO, ML, CCR, and EJA. Data acquisition was coordinated by APA, AZ, DO, ML, RZP, SRRC, and EJA. Laboratory analysis and interpretation was coordinated by MRL and CCR. Data analysis and report generation was performed by ML, KLC, DO, and EJA. Mapping was done by KLC. Preparation of the manuscript was led by DO, ML, and EJA. All authors reviewed and approved the final version.

Conflicts of Interest

Dr Asturias has served on an Advisory Board for Takeda Vaccines Inc and is partially supported by research grants from GlaxoSmithKline Biologicals and Takeda Vaccines Inc. Dr Lamb is partially supported by grants from GlaxoSmithKline Biologicals and Pantheryx Inc. Dr Olson is partially supported by a grant from Takeda Vaccines Inc. Ricardo Zambrano-Perilla and Sergio Ricardo Rodríguez-Castro are employed by Integra IT and own stake in the company.

Multimedia Appendix 1

Sample screenshots from the Vigilant-e application (Integra IT, Bogota, Colombia) for weekly participatory syndromic surveillance in the Trifinio Region of Guatemala. Each week, subjects would select if any of the children enrolled from their household had fever or rash (Panel A). If the parent or guardian reports symptoms (red box), a decision-tree logic would then ask them to report symptoms for each child (Panel B). If they selected the child and then reported fever, vomiting, or diarrhea (not shown), another screen would appear asking additional syndromic questions (Panel C).

[JPG File, 64KB - jmir_v19i11e368_app1.JPG]

References

1. Li Y, Fang L, Gao S, Wang Z, Gao H, Liu P, et al. Decision support system for the response to infectious disease emergencies based on WebGIS and mobile services in China. PLoS One 2013;8(1):e54842 [FREE Full text] [doi: 10.1371/journal.pone.0054842] [Medline: 23372780]
2. Robertson C, Sawford K, Daniel SL, Nelson TA, Stephen C. Mobile phone-based infectious disease surveillance system, Sri Lanka. Emerg Infect Dis 2010 Oct;16(10):1524-1531 [FREE Full text] [doi: 10.3201/eid1610.100249] [Medline: 20875276]
3. Lin Y, Heffernan C. Accessible and inexpensive tools for global HPAI surveillance: a mobile-phone based system. Prev Vet Med 2011 Feb 01;98(2-3):209-214. [doi: 10.1016/j.prevetmed.2010.10.003] [Medline: 21167614]
4. Yang C, Yang J, Luo X, Gong P. Use of mobile phones in an emergency reporting system for infectious disease surveillance after the Sichuan earthquake in China. Bull World Health Organ 2009 Aug;87(8):619-623 [FREE Full text] [Medline: 19705013]
5. Agarwal S, Perry HB, Long L, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
6. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, Guidelines Working Group Centers for Disease Control and Prevention. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep 2001 Jul 27;50(RR-13):1-35; quiz CE1. [Medline: 18634202]
7. Sarti E, L'Azou M, Mercado M, Kuri P, Siqueira JB, Solis E, et al. A comparative study on active and passive epidemiological surveillance for dengue in five countries of Latin America. Int J Infect Dis 2016 Mar;44:44-49 [FREE Full text] [doi: 10.1016/j.ijid.2016.01.015] [Medline: 26836763]
8. Global Health Workforce Alliance, World Health Organization. Global Health Workforce Alliance. 2015. A universal truth: no health without a workforce. Third Global Forum on Human Resources for Health report URL: http://www.who.int/workforcealliance/knowledge/resources/hrhreport2013/en/ [accessed 2017-10-11] [WebCite Cache ID 6u92ZNtUa]
9. Wójcik OP, Brownstein JS, Chunara R, Johansson MA. Public health for the people: participatory infectious disease surveillance in the digital age. Emerg Themes Epidemiol 2014;11:7 [FREE Full text] [doi: 10.1186/1742-7622-11-7] [Medline: 24991229]
10. Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004;11(2):141-150 [FREE Full text] [doi: 10.1197/jamia.M1356] [Medline: 14633933]
11. Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168. [doi: 10.1016/S1473-3099(13)70244-5] [Medline: 24290841]

12. Pini A, Merk H, Carnahan A, Galanis I, Danis K, Edelstein M, et al. High added value of a population-based participatory surveillance system for community acute gastrointestinal, respiratory and influenza-like illnesses in Sweden, 2013-2014 using the web. Epidemiol Infect 2017 Apr;145(6):1193-1202. [doi: 10.1017/S0950268816003290] [Medline: 28137317]
13. Lester J, Paige S, Chapman CA, Gibson M, Holland JJ, Switzer WM, et al. Assessing commitment and reporting fidelity to a text message-based participatory surveillance in rural western Uganda. PLoS One 2016;11(6):e0155971 [FREE Full text] [doi: 10.1371/journal.pone.0155971] [Medline: 27281020]
14. Chenais E, Sternberg-Lewerin S, Boqvist S, Emanuelson U, Aliro T, Tejler E, et al. African swine fever in Uganda: qualitative evaluation of three surveillance methods with implications for other resource-poor settings. Front Vet Sci 2015;2:51 [FREE Full text] [doi: 10.3389/fvets.2015.00051] [Medline: 26664978]
15. White A, Thomas DS, Ezeanochie N, Bull S. Health worker mHealth utilization: a systematic review. Comput Inform Nurs 2016 Mar 7:1. [doi: 10.1097/CIN.0000000000000231] [Medline: 26955009]
16. Situación Epidemiológica de Enfermedades Transmitidas por Agua y Alimentos (ETAs): Año 2017, S. E 1-27 Epidemiología. Guatemala City: Departamento de Epidemiología, Ministerio de Salud Pública y Asistencia Social; 2017. URL: http://epidemiologia.mspas.gob.gt/files/Publicaciones%202017/ETAS/ETAS_SEM_27_2017.pdf [accessed 2017-10-11] [WebCite Cache ID 6u92WoPoG]
17. Ministerio de Salud Pública y Asistencia Social. Semana Epidemiológica 2017. 2017 Apr 17. URL: http://epidemiologia.mspas.gob.gt/files/Publicaciones%202017/SEMEPI%202017/SEMEPI_14_2017.pdf [accessed 2017-10-11] [WebCite Cache ID 6u92leokF]
18. Lopman B. Global Burden of Norovirus and Prospects for Vaccine Development. Atlanta, GA: Centers for Disease Control and Prevention; 2015. URL: https://www.cdc.gov/norovirus/downloads/global-burden-report.pdf [accessed 2017-10-11] [WebCite Cache ID 6u92Ugb7B]
19. L'Azou M, Moureau A, Sarti E, Nealon J, Zambrano B, Wartel TA, CYD14 Primary Study Group, CYD15 Primary Study Group. Symptomatic dengue in children in 10 Asian and Latin American countries. N Engl J Med 2016 Mar 24;374(12):1155-1166. [doi: 10.1056/NEJMoa1503877] [Medline: 27007959]
20. Asturias EJ, Heinrichs G, Domek G, Brett J, Shick E, Cunningham M, et al. The Center for Human Development in Guatemala: an innovative model for global population health. Adv Pediatr 2016 Aug;63(1):357-387. [doi: 10.1016/j.yapd.2016.04.001] [Medline: 27426907]
21. Pocket Book of Hospital Care for Children: Guidelines for the Management of Common Childhood Illnesses. Geneva: World Health Organization; 2013.
22. Singh J, Jain DC, Sharma RS, Verghese T. Evaluation of immunization coverage by lot quality assurance sampling compared with 30-cluster sampling in a primary health centre in India. Bull World Health Organ 1996;74(3):269-274 [FREE Full text] [Medline: 8789925]
23. D'Ardenne KK, Darrow J, Furniss A, Chavez C, Hernandez H, Berman S, et al. Use of rapid needs assessment as a tool to identify vaccination delays in Guatemala and Peru. Vaccine 2016 Mar 29;34(14):1719-1725. [doi: 10.1016/j.vaccine.2016.01.060] [Medline: 26902545]
24. Lanata CF, Black RE. Lot quality assurance sampling techniques in health surveys in developing countries: advantages and current constraints. World Health Stat Q 1991;44(3):133-139. [Medline: 1949880]
25. Olson D, Lamb M, Lopez M, Paniagua-Avila M, Zacarias A, Samayoa-Reyes G, et al. A rapid epidemiological tool to measure the burden of norovirus infection and disease in resource-limited settings. Open Forum Infect Dis 2017;4(2):ofx049 [FREE Full text] [doi: 10.1093/ofid/ofx049] [Medline: 28730158]
26. Goldfarb DM, Steenhoff AP, Pernica JM, Chong S, Luinstra K, Mokomane M, et al. Evaluation of anatomically designed flocked rectal swabs for molecular detection of enteric pathogens in children admitted to hospital with severe gastroenteritis in Botswana. J Clin Microbiol 2014 Nov;52(11):3922-3927 [FREE Full text] [doi: 10.1128/JCM.01894-14] [Medline: 25165077]
27. Arvelo W, Hall AJ, Estevez A, Lopez B, Gregoricus N, Vinjé J, et al. Diagnostic performance of rectal swab versus bulk stool specimens for the detection of rotavirus and norovirus: implications for outbreak investigations. J Clin Virol 2013 Dec;58(4):678-682 [FREE Full text] [doi: 10.1016/j.jcv.2013.09.019] [Medline: 24139675]
28. Trujillo AA, McCaustland KA, Zheng D, Hadley LA, Vaughn G, Adams SM, et al. Use of TaqMan real-time reverse transcription-PCR for rapid detection, quantification, and typing of norovirus. J Clin Microbiol 2006 Apr;44(4):1405-1412 [FREE Full text] [doi: 10.1128/JCM.44.4.1405-1412.2006] [Medline: 16597869]
29. Debin M, Turbelin C, Blanchon T, Bonmarin I, Falchi A, Hanslik T, et al. Evaluating the feasibility and participants' representativeness of an online nationwide surveillance system for influenza in France. PLoS One 2013;8(9):e73675 [FREE Full text] [doi: 10.1371/journal.pone.0073675] [Medline: 24040020]
30. Dalton C, Carlson S, Butler M, Cassano D, Clarke S, Fejsa J, et al. Insights from Flutracking: thirteen tips to growing a Web-based participatory surveillance system. JMIR Public Health Surveill 2017 Aug 17;3(3):e48 [FREE Full text] [doi: 10.2196/publichealth.7333] [Medline: 28818817]
31. R: A language and environment for statistical computing. Vienna, Austria: R Development Core Team; 2011.
32. Rouhani S, Peñataro YP, Paredes OM, Siguas SM, Rengifo TD, Mondal D, Etiology, Risk Factors, and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development Project (MAL-ED) Network

Investigators. Norovirus infection and acquired immunity in 8 countries: results from the MAL-ED study. Clin Infect Dis 2016 May 15;62(10):1210-1217 [FREE Full text] [doi: 10.1093/cid/ciw072] [Medline: 27013692]

33. Fischer WC, Perin J, Aryee MJ, Boschi-Pinto C, Black RE. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. BMC Public Health 2012 Mar 21;12:220 [FREE Full text] [doi: 10.1186/1471-2458-12-220] [Medline: 22436130]

Abbreviations

ELISA: enzyme-linked immunosorbent assay
HCW: health care workers
IMCI: Integrated Management of Childhood Illness
LMIC: low- and middle-income countries
N/C: not calculated
RR: risk ratio
RT-PCR: reverse transcription polymerase chain reaction

Edited by G Eysenbach; submitted 12.05.17; peer-reviewed by Y Jo, C Fu; comments to author 30.08.17; revised version received 13.09.17; accepted 13.09.17; published 09.11.17.

Please cite as:
Olson D, Lamb M, Lopez MR, Colborn K, Paniagua-Avila A, Zacarias A, Zambrano-Perilla R, Rodríguez-Castro SR, Cordon-Rosales C, Asturias EJ
Performance of a Mobile Phone App-Based Participatory Syndromic Surveillance System for Acute Febrile Illness and Acute Gastroenteritis in Rural Guatemala
J Med Internet Res 2017;19(11):e368
URL: http://www.jmir.org/2017/11/e368/
doi:10.2196/jmir.8041
PMID:29122738

Original Paper

Artificial Intelligence Learning Semantics via External Resources for Classifying Diagnosis Codes in Discharge Notes

Chin Lin[1,2], PhD; Chia-Jung Hsu[3], BS; Yu-Sheng Lou[1], BS; Shih-Jen Yeh[4], PhD; Chia-Cheng Lee[3*], MD; Sui-Lung Su[1*], PhD; Hsiang-Cheng Chen[5*], MD, PhD

[1]School of Public Health, National Defense Medical Center, Taipei, Taiwan

[2]Department of Research and Development, National Defense Medical Center, Taipei, Taiwan

[3]Planning and Management Office, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[4]Da-Yeh University, Changhua, Taiwan

[5]Division of Rheumatology/Immunology/Allergy, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Hsiang-Cheng Chen, MD, PhD
Division of Rheumatology/Immunology/Allergy, Department of Internal Medicine
Tri-Service General Hospital
National Defense Medical Center
No.161, Min-Chun E. Rd., Sec. 6, Neihu
Taipei, 114
Taiwan
Phone: 886 2 7927135
Email: hccheng@ndmctsgh.edu.tw

Abstract

Background: Automated disease code classification using free-text medical information is important for public health surveillance. However, traditional natural language processing (NLP) pipelines are limited, so we propose a method combining word embedding with a convolutional neural network (CNN).

Objective: Our objective was to compare the performance of traditional pipelines (NLP plus supervised machine learning models) with that of word embedding combined with a CNN in conducting a classification task identifying *International Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*) diagnosis codes in discharge notes.

Methods: We used 2 classification methods: (1) extracting from discharge notes some features (terms, n-gram phrases, and SNOMED CT categories) that we used to train a set of supervised machine learning models (support vector machine, random forests, and gradient boosting machine), and (2) building a feature matrix, by a pretrained word embedding model, that we used to train a CNN. We used these methods to identify the chapter-level *ICD-10-CM* diagnosis codes in a set of discharge notes. We conducted the evaluation using 103,390 discharge notes covering patients hospitalized from June 1, 2015 to January 31, 2017 in the Tri-Service General Hospital in Taipei, Taiwan. We used the receiver operating characteristic curve as an evaluation measure, and calculated the area under the curve (AUC) and F-measure as the global measure of effectiveness.

Results: In 5-fold cross-validation tests, our method had a higher testing accuracy (mean AUC 0.9696; mean F-measure 0.9086) than traditional NLP-based approaches (mean AUC range 0.8183-0.9571; mean F-measure range 0.5050-0.8739). A real-world simulation that split the training sample and the testing sample by date verified this result (mean AUC 0.9645; mean F-measure 0.9003 using the proposed method). Further analysis showed that the convolutional layers of the CNN effectively identified a large number of keywords and automatically extracted enough concepts to predict the diagnosis codes.

Conclusions: Word embedding combined with a CNN showed outstanding performance compared with traditional methods, needing very little data preprocessing. This shows that future studies will not be limited by incomplete dictionaries. A large amount of unstructured information from free-text medical writing will be extracted by automated approaches in the future, and we believe that the health care field is about to enter the age of big data.

(J Med Internet Res 2017;19(11):e380) doi:10.2196/jmir.8344

KEYWORDS

word embedding; convolutional neural network; neural networks (computer); natural language processing; text mining; data mining; machine learning; electronic medical records; electronic health records

Introduction

Public health surveillance systems are important for identifying unusual events of public health importance and will provide information for public health action [1]. However, most surveillance systems can only use structured data, such as *International Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*) diagnosis codes. The current methods for collecting this structured information usually involve manual identification, but manual identification of disease codes from free-text clinical narratives is laborious and costly. Moreover, most surveillance systems do not have enough expert clinical coders for real-time surveillance, and this leads to delays in the release of disease statistics. Government health administrators need timely information to rapidly assess disease prevention and health protection priorities. A timely and computer-based disease classification approach is required to further assist public health action.

Automated surveillance methods are increasingly being researched because of the increasing volume and accessibility of electronic medical data, and a range of studies have proven the feasibility of extracting structured information from clinical narratives [2-6]. Previous studies suggested that these text mining approaches would need to effectively deal with the idiosyncrasies of the clinical sublanguage to further improve performance [7]. However, compiling a complete medical dictionary may be impossible because of the variability of clinical vocabularies. Moreover, traditional natural language processing (NLP) pipelines can deal with synonyms but not similar terms, so supervised machine learning models often face the curse of dimensionality. For example, if we only want to identify infectious disease-related medical documents, bacteria names such as *Streptococcus pneumoniae* and *Mycobacterium tuberculosis* can actually be treated as similar for classification purposes. An effective text preprocessing approach would need to learn how to combine similar concepts, and current NLP pipelines often cannot deal with this issue.

Another important challenge for automated surveillance algorithms is emerging disease. For example, influenza H1N1 broke out in 2009 and could not have been recorded in any medical records before 2008. Traditional automatic methods based on term vectors cannot use new terms [2-6]. This weakness means that traditional methods cannot possibly implement a fully automated pipeline. The key reason that human experts can successfully identify emerging diseases is that humans can learn semantics from external resources. Traditionally, these external resources usually take the form of a dictionary, and this is what will be used in the NLP pipeline. However, dictionary construction is laborious, and it is still difficult to completely include all semantic relationships. In summary, traditional NLP pipelines are complex and inefficient, and successful automated surveillance methods will also need to include automatic handling of semantics.

Word embedding is a feature learning technique where vocabularies are mapped to vectors of real numbers [8,9]. Word2vec [10] and GloVe [11] are the 2 most popular word embedding algorithms. These methods showcase interesting linear substructures in the word vector space: word vectors for similar concepts are likewise close in terms of cosine similarity and Euclidean distance. This property may help us identify concept groups and reduce the data dimensionality in future machine learning algorithms. However, clinical narratives will be transformed into a matrix, and standardization to vectors with different length is difficult for general machine learning models. Convolutional neural networks (CNNs) use layers with convolving filters that are applied to local features, and they can handle matrix input [12]. CNNs were originally invented for computer vision applications and have subsequently been shown to achieve excellent results in semantic parsing [13], search query retrieval [14], and sentence classification [15]. The key reason for the success of CNNs is their fuzzy matching using convolving filters, and we believe that convolving filters are a great way to process similar texts involving the same concepts. A lot of words and phrases that are conceptually similar can be combined in a convolving filter via fuzzy matching technology, thereby reducing the data dimensionality and avoiding overfitting.

This project aimed to compare traditional machine learning pipelines (NLP plus supervised machine learning models) versus word embedding combined with a CNN in order to identify chapter-level *ICD-10-CM* diagnosis codes in discharge notes. We hoped to develop an efficient and effective real-time surveillance pipeline for disease statistics. In addition, we further analyzed the convolving filters of the CNN to understand their functions.

Methods

Data Source

The Tri-Service General Hospital, Taipei, Taiwan, supplied deidentified free-text discharge notes from June 1, 2015 to January 31, 2017. Research ethics approval was given by the institutional ethical committee and medical records office of the Tri-Service General Hospital to collect data without individual consent for sites where data are directly collected. The Tri-Service General Hospital is located in the Neihu District of Taipei under the name of National Defense Medical Center and provides medical service for service members, their family dependents, and civilians. It has been rated by the Ministry of Health and Welfare in Taiwan as a first-rate teaching hospital on the level of a medical center. The hospital has about 1700 beds and 6000 inpatients per month, and most inpatients are civilians. We collected a total of 103,390 discharge notes, and corrected misspellings using the Hunspell version 2.3 package [16] and a dictionary built using English Wikipedia and Gigaword [17]. *ICD-10-CM* codes had been used to label these discharge notes for the purpose of requesting health insurance

fees, and the medical records department was responsible for their correctness. The Taiwan National Health Insurance Administration routinely samples a certain number of discharge notes for verification, and a wrongly labeled discharge note is punishable by a 10- to 20-fold fine. Discharge notes are often labeled with multiple *ICD-10-CM* codes, and all *ICD-10-CM* codes were truncated at the 1-character level. There are a total of 21 categories in the 2017 version. Table 1 shows the frequency distribution of 1-character-level codes. Neoplasms and diseases of the circulatory system were the most common *ICD-10-CM* codes in our hospital.

We used 2 testing procedures to assess the performance of the model. First, we conducted a 5-fold cross-validation test. Second, we created training and testing sets by splitting the sample by date (July 1, 2016), because this is more realistic. A classifier can only be trained using retrospective data in the real world, and it will be used to classify future data; the second testing process replicates this. All calculations were conducted on a Fujitsu RX2540M1 48-core CPU, 768 GB RAM server (Fujitsu Ltd, Tokyo, Japan), and the all-flash array was AccelStor NeoSapphire NS3505 (AccelStor, Inc, Taipei City, Taiwan) with a 5 TB serial advanced technology attachment-interface solid-state drive and connectivity of 56 GB/second FDR InfiniBand Quad Small Form-factor Pluggable (Fiberon Technologies, Inc, Westborough, MA, USA).

Table 1. Prevalence of different *International Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*) chapter-level codes in discharge notes from the Tri-Service General Hospital, Taipei, Taiwan.

ICD-10-CM code	Definition	Stage of the study		
		Before June 30, 2016 (n=64,023) n (%)	After July 1, 2016 (n=39,367) n (%)	Full study period (n=103,390) n (%)
A00-B99	Certain infectious and parasitic diseases	7731 (12.1%)	5455 (13.9%)	13,186 (12.8%)
C00-D49	Neoplasms	20,585 (32.2%)	13,993 (35.5%)	34,578 (33.5%)
D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	4516 (7.1%)	3132 (8.0%)	7648 (7.4%)
E00-E89	Endocrine, nutritional, and metabolic diseases	13,223 (20.7%)	8765 (22.3%)	21,988 (21.3%)
F01-F99	Mental, behavioral, and neurodevelopmental disorders	4612 (7.2%)	2942 (7.5%)	7554 (7.3%)
G00-G99	Diseases of the nervous system	3703 (5.8%)	2602 (6.6%)	6305 (6.1%)
H00-H59	Diseases of the eye and adnexa	2337 (3.7%)	1374 (3.5%)	3711 (3.6%)
H60-H95	Diseases of the ear and mastoid process	802 (1.3%)	470 (1.2%)	1272 (1.2%)
I00-I99	Diseases of the circulatory system	17,650 (27.6%)	11,465 (29.1%)	29,115 (28.2%)
J00-J99	Diseases of the respiratory system	7743 (12.1%)	5584 (14.2%)	13,327 (13.0%)
K00-K95	Diseases of the digestive system	12,849 (20.1%)	8444 (21.4%)	21,293 (20.6%)
L00-L99	Diseases of the skin and subcutaneous tissue	2568 (4.0%)	1711 (4.3%)	4279 (4.1%)
M00-M99	Diseases of the musculoskeletal system and connective tissue	9170 (14.3%)	5152 (13.1%)	14,322 (13.9%)
N00-N99	Diseases of the genitourinary system	9929 (15.5%)	7325 (18.6%)	17,254 (16.8%)
O00-O9A	Pregnancy, childbirth, and the puerperium	2509 (3.9%)	1271 (3.2%)	3780 (3.7%)
P00-P96	Certain conditions originating in the perinatal period	793 (1.2%)	493 (1.3%)	1286 (1.2%)
Q00-Q99	Congenital malformations, deformations, and chromosomal abnormalities	927 (1.4%)	513 (1.3%)	1440 (1.4%)
R00-R99	Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	5271 (8.2%)	3824 (9.7%)	9095 (8.9%)
S00-T88	Injury, poisoning, and certain other consequences of external causes	6272 (9.8%)	4564 (11.6%)	10,836 (10.6%)
V00-Y99	External causes of morbidity	791 (1.2%)	68 (0.2%)	859 (0.8%)
Z00-Z99	Factors influencing health status and contact with health services	15,488 (24.2%)	10,093 (25.6%)	25,581 (24.8%)

Traditional Free-Text Classification Techniques

Traditional classification techniques often combine an NLP pipeline and a classifier to conduct free-text medical writing classification tasks. We extracted the detailed features from the discharge notes by the NLP pipeline; then *ICD-10-CM* codes were assigned by human experts to each discharge note. We used the labeled features to train a classifier, and we used the well-trained model to predict the unlabeled testing data.

In this study, we used a 2-part NLP pipeline to extract the discharge note features. First, word-based features were directly extracted from the free-text description and n-gram phrases (n range 2-5) were generated by the RWeka version 0.4-30 package [18]. To reduce the complexity of the data, we only included n-gram phrases with counts >10. Second, we used SNOMED CT International Edition version 20170131 categories to integrate synonyms. We used the bag-of-words model to vectorize the extracted features (1 vector per discharge note) and transformed these feature vectors into a document-term matrix using the tm version 0.7 package [19]. This matrix was then the input into the following machine learning models.

Support Vector Machine

Support vector machines (SVMs) are common classifiers in the machine learning field. They map all samples onto a hyperplane and divide them by a clear gap. In addition, kernel tricks are used to extend this hyperplane. SVMs are proven to have the best performance in free-text medical writing classification, compared with naive Bayes classifiers, C4.5 decision trees, and adaptive boosting [20]. In this study, we used the 4 most common kernel tricks: linear, polynomial (degree=3), radial basis, and sigmoid. We used the e1071 package (R package version 1.6-8) [21] as the SVM implementation and set all other parameters to their default values.

Random Forest

Random forests (RFs) construct multiple decision trees and use information from each tree to make predictions. It was the best-performing classification model in a previous text classification study [22], compared with SVMs, naive Bayes classifiers, and the k-nearest neighbors algorithm. We used the H2O version 3.10.2.2 package [23] as the RF implementation and set all parameters to their default values.

Gradient Boosting Machine

Gradient boosting machines (GBMs) are also ensembles of weak decision trees, where the gradient boosting method is used to improve the predictive ability of each tree [24]. They use greedy function approximation to build a series of weak trees [25]. The H2O package also provides the function for the GBM implementation, and we set all parameters to their default values.

Using the “no free lunch” theorem [26], we combined a traditional NLP pipeline with the 3 abovementioned models and tested their performance on our task.

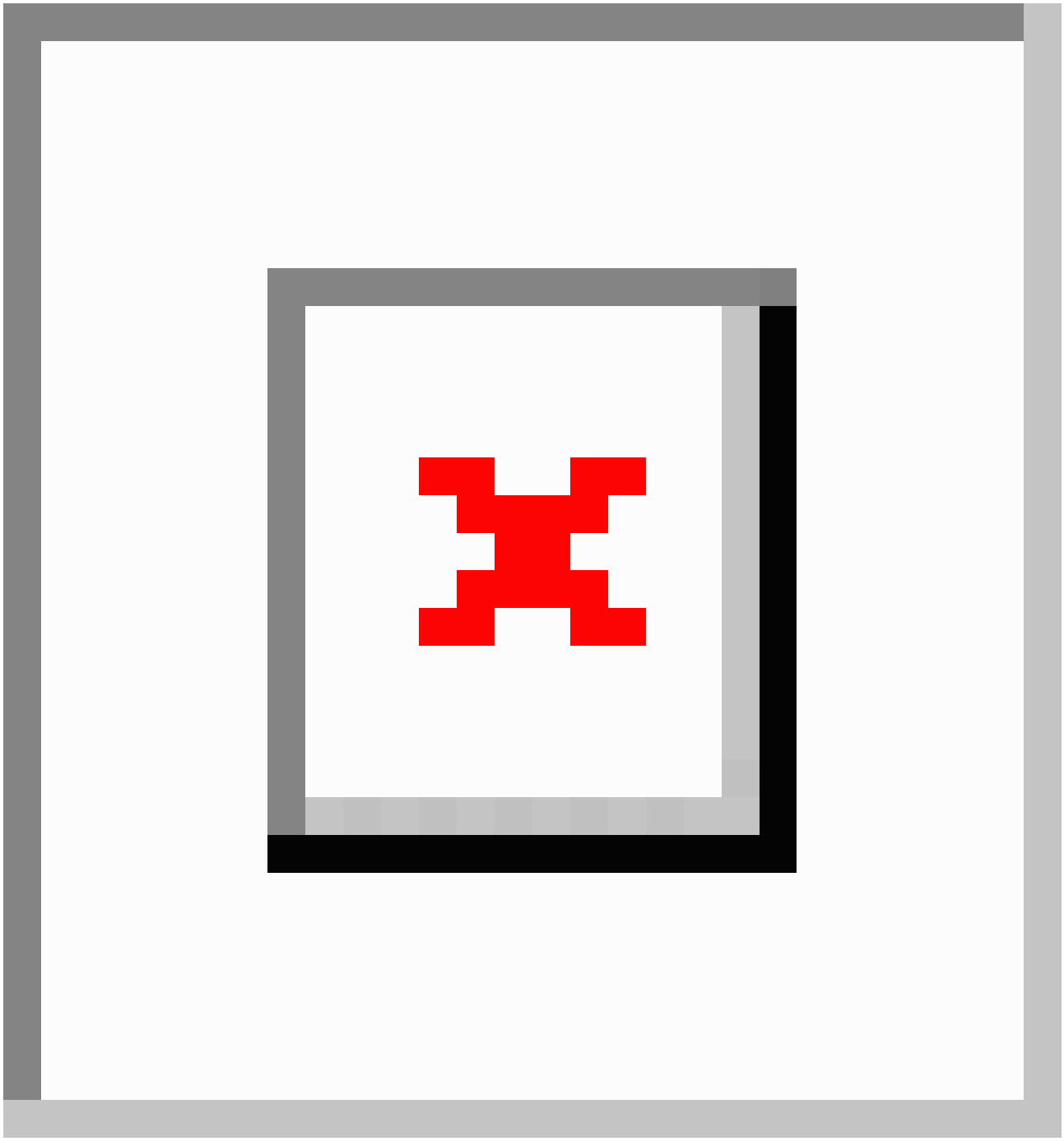
Word Embedding Combined With a Convolutional Neural Network

Traditional NLP pipelines are limited by their preexisting dictionary and need to build a complex processing flow. Herein, we propose a method combining a word embedding model and a CNN. Word embedding technology is useful for integrating synonyms, and we used a pretrained GloVe model (English Wikipedia plus Gigaword) to vectorize the words. We selected a 50-dimensional model with 400,000 words because of computing time constraints. However, we believe that this was sufficient because there were only 19,064 words in our 103,390 discharge notes. We transformed each discharge note into an n×50 matrix for subsequent classification (where n is the number of words in the discharge note) and trained a CNN using these labeled matrixes.

Although CNNs with various structures have been developed, we focused on a 1-layer CNN with a filter region size of 1-5 (corresponding to 1-5 n-gram phrases) to increase comparability with traditional machine learning technologies. In fact, these simple models have recently achieved remarkably strong performance [15,27,28]. Figure 1 shows the proposed model’s architecture. We set 5 convolution channels, and their convolution layers were as follows: (1) 40 convolving filters with a 1×50 region size, to identify the important words; (2) 30 convolving filters with a 2×50 region size, to identify the important 2-gram phrases; (3) 15 convolving filters with a 3×50 region size, to identify the important 3-gram phrases; (4) 10 convolving filters with a 4×50 region size, to identify the important 4-gram phrases; and (5) 5 convolving filters with a 5×50 region size, to identify the important 5-gram phrases. These convolution layers were connected to a rectified linear unit layer to enhance the nonlinearity of the network. We then applied a max pooling layer over the feature map and took the maximum value. The above steps are similar to those of the keyword recognition process, and 100 features were extracted from each discharge note. To avoid the risk of overfitting, we used a dropout layer with a 50% drop rate after the convolution channels [29]. Finally, we used logistic regression to connect the features, and the cross-entropy loss function in the loss layer to train the CNN.

We used the MXNet version 0.8.0 package [30] to implement the above architecture. The settings used for the training model were as follows: (1) minibatch gradient descent with 1000 bench size for optimization; (2) learning rate=.05; (3) momentum coefficient=.9; (4) L2 regularization coefficient=.00001; and (5) tolerance of early stopping per 100 iterations=.0001. Multimedia Appendix 1 shows an example code for implementing the word embedding and CNNs for free-text discharge note classification.

Figure 1. Model architecture with 5 convolution channels and 1 full connection (FC) layer. ReLU: rectified linear unit.



Model Details and Evaluation Index

We conducted oversampling processing for sufficiently regarding positive cases but not skewing by an overwhelming number of negative cases [31,32]. All the models return a continuous value to evaluate model performance. SVM-related models provide the decision values of the binary classifier; RF and GBM models provide the mean of the probabilities from the decision trees; and CNNs provide the probabilities calculated by the logistic function. We used the receiver operating characteristic curve as an evaluation measure, and the area under the curve (AUC) provided a global measure of effectiveness. Moreover, we provide the F-measure, which is calculated by following equations: *precision* = *TruePositives* / (*TruePositives* + *FalsePositives*); *recall* = *TruePositives* / (*TruePositives* + *FalseNegatives*); *F-measure* = (2 × *precision* × *recall*) / (*precision* + *recall*).

Results

Cross-Validation Test

Table 2 shows the global and lowest 5 means of the training and testing AUCs in the 5-fold cross-validation test. The proposed word embedding plus CNN method provided the highest AUCs (mean testing AUC = 0.9696; mean of the lowest 5 AUCs = 0.9135) and highest F-measures (mean testing F-measure = 0.9086; mean of the lowest 5 F-measures = 0.7651). It is worth noting that the SVM with the linear kernel trick had

the highest mean testing AUC of all the traditional methods (mean testing AUC = 0.9571; mean of the lowest 5 AUCs = 0.8891). The performances of the RF, GBM, and linear SVM models were similar (mean testing AUCs of 0.9570, 0.9544, and 0.9571, respectively). However, the RF and GBM models were very inefficient in some tasks (as Multimedia Appendix 2 shows). The RF and GBM models had a lower mean testing AUC owing to the V00-Y99 *ICD-10-CM* code identification tasks; therefore, the linear SVM was a relatively stable model.

Table 2. Global (and lowest 5) means of training and testing AUCs[a] in the 5-fold cross-validation test.

Pipeline	Training set		Testing set	
	AUC[b]	F-measure	AUC[b]	F-measure
Traditional				
NLP[c] + SVM[d] (linear)	0.9947 (0.9836)	0.9546 (0.8560)	0.9571 (0.8891)	0.8606 (0.6387)
NLP + SVM (polynomial)	0.8627 (0.6736)	0.5630 (0.2498)	0.8183 (0.6332)	0.5050 (0.2023)
NLP + SVM (radial basis)	0.9565 (0.9146)	0.7984 (0.6613)	0.9363 (0.8582)	0.7569 (0.5352)
NLP + SVM (sigmoid)	0.9518 (0.9021)	0.7852 (0.6368)	0.9325 (0.8526)	0.7498 (0.5313)
NLP + RF[e]	0.9999 (0.9995)[f]	0.9864 (0.9628)	0.9570 (0.8800)	0.8739 (0.6475)
NLP + GBM[g]	0.9996 (0.9990)	0.9868 (0.9660)	0.9544 (0.8722)	0.8691 (0.6458)
Proposed				
GloVe[h] + CNN[i]	0.9964 (0.9890)	0.9837 (0.9588)	0.9696 (0.9135)[f]	0.9086 (0.7651)

[a]AUC: area under the curve, calculated using the receiver operating characteristic curve.

[b]The results are presented as the mean AUC or F-measure (mean of the lowest 5 AUCs or F-measures). Detailed AUCs and F-measures for each chapter-level *International Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*) diagnosis code are shown in Multimedia Appendix 2.

[c]NLP: natural language processing for feature extraction (terms, n-gram phrases, and SNOMED CT categories).

[d]SVM: support vector machine.

[e]RF: random forest.

[f]The best method for a specific index.

[g]GBM: gradient boosting machine.

[h]GloVe: a 50-dimensional word embedding model, pretrained using English Wikipedia and Gigaword.

[i]CNN: convolutional neural network.

Real-World Test

Table 3 shows the global and lowest 5 means of the training and testing AUCs in the real-world test, where the testing samples were split by date. The results of this test were similar to those of the cross-validation test. The testing AUC in the real-world test was lower than that in the cross-validation test, possibly because the heterogeneity between the training and testing samples was higher in the real-world test owing to there being many cyclical diseases. However, our proposed method still had the highest performance on the testing set (mean testing AUC = 0.9645; mean testing F-measure = 0.9003; mean of the lowest 5 AUCs = 0.8952; mean of the lowest 5 F-measures = 0.7204) and achieved the best results in almost all tasks. Multimedia Appendix 3 shows the detailed training and testing AUCs. The testing AUC of the proposed method is only obviously worse than that of traditional methods for the Q00-Q99 code identification tasks. In addition, the performances of all methods were bad for the V00-Y99 code identification tasks.

Convolving Filter Analysis

We visualized 3 of the convolving filters selected for the real-world test, as Figure 2 shows. Neoplasms were the most common *ICD-10-CM* codes in our hospital, and we selected the filter with highest information gain for these. Information gain can be estimated as $IG(C, F) = H(C) - H(C \mid F)$, where C is the class (a specific *ICD-10-CM* code), F is the feature extracted by the convolving filter, and H is the information entropy function. This filter is a word filter that identified several cancer-related words, such as carcinoma and adenocarcinoma, when trained using the training data (Figure 2, panel A). As expected, these words, embodying similar concepts, were identified by the fuzzy matching technology. Moreover, the same words in the testing data were identified by this convolving filter (Figure 2, panel B). Figure 2, panel C shows a 2-gram convolving filter for certain infectious and parasitic diseases, which can identify many pathogens. It is worth mentioning that some pathogens absent in the training data were identified by this filter in the testing data (Figure 2, panel D). Identifying the external causes of morbidity was the most difficult task for all of the methods, and Figure 2, panel E shows the most important filter for this task. Some accident-related words were identified, such as fracture and injury, but these words were widely used in our discharge notes. The total number of discharge notes that included these words was 7855, but only 791 discharge notes were coded as V00-Y99 in the training set. This caused the

information gain to be very low for the testing set (Figure 2, panel F).

Figure 3 shows the information gain distribution of the convolving filters in each task, demonstrating large differences between them. The highest-performing classification tasks often extracted high information gain features using convolving filters. Moreover, when the geometric mean of the information gain ratio between the training and testing sets was over 80%, the testing AUC was more than 0.98. It is worth noting that the information gain ratio was very low for Q00-Q99 and V00-Y99 (19.9% and 0.9%, respectively). This may explain the lower performance in these tasks.

Table 3. Global (and lowest 5) means of the training and testing AUCs[a] in the real-world test.

Pipeline	Training set		Testing set	
	AUC[b]	F-measure	AUC[b]	F-measure
Traditional				
NLP[c] + SVM[d] (linear)	0.9921 (0.9768)	0.9365 (0.7983)	0.9477 (0.8549)	0.8458 (0.5984)
NLP + SVM (polynomial)	0.9103 (0.7975)	0.6316 (0.4045)	0.8716 (0.7400)	0.5761 (0.2802)
NLP + SVM (radial basis)	0.9577 (0.9208)	0.7954 (0.6484)	0.9349 (0.8476)	0.7588 (0.5258)
NLP + SVM (sigmoid)	0.9522 (0.9058)	0.7840 (0.6261)	0.9259 (0.8196)	0.7515 (0.5209)
NLP + RF[e]	0.9996 (0.9985)[f]	0.9869 (0.9664)[f]	0.9483 (0.8484)	0.8582 (0.5901)
NLP + GBM[g]	0.9995 (0.9985)	0.9821 (0.9562)	0.9462 (0.8416)	0.8568 (0.5948)
Proposed				
GloVe[h] + CNN[i]	0.9956 (0.9868)	0.9803 (0.9523)	0.9645 (0.8952)[f]	0.9003 (0.7204)[f]

[a]AUC: area under the curve, calculated using the receiver operating characteristic curve.

[b]The results are presented as the mean AUC or F-measure (mean of the lowest 5 AUCs or F-measures). Detailed AUCs and F-measures for each chapter-level *International Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*) diagnosis code are shown in Multimedia Appendix 3.

[c]NLP: natural language processing for feature extraction (terms, n-gram phrases, and SNOMED CT categories).

[d]SVM: support vector machine.

[e]RF: random forest.

[f]The best method for a specific index.

[g]GBM: gradient boosting machine.

[h]GloVe: a 50-dimensional word embedding model, pretrained using English Wikipedia and Gigaword.

[i]CNN: convolutional neural network.

Figure 2. Visualization of selected convolving filters.

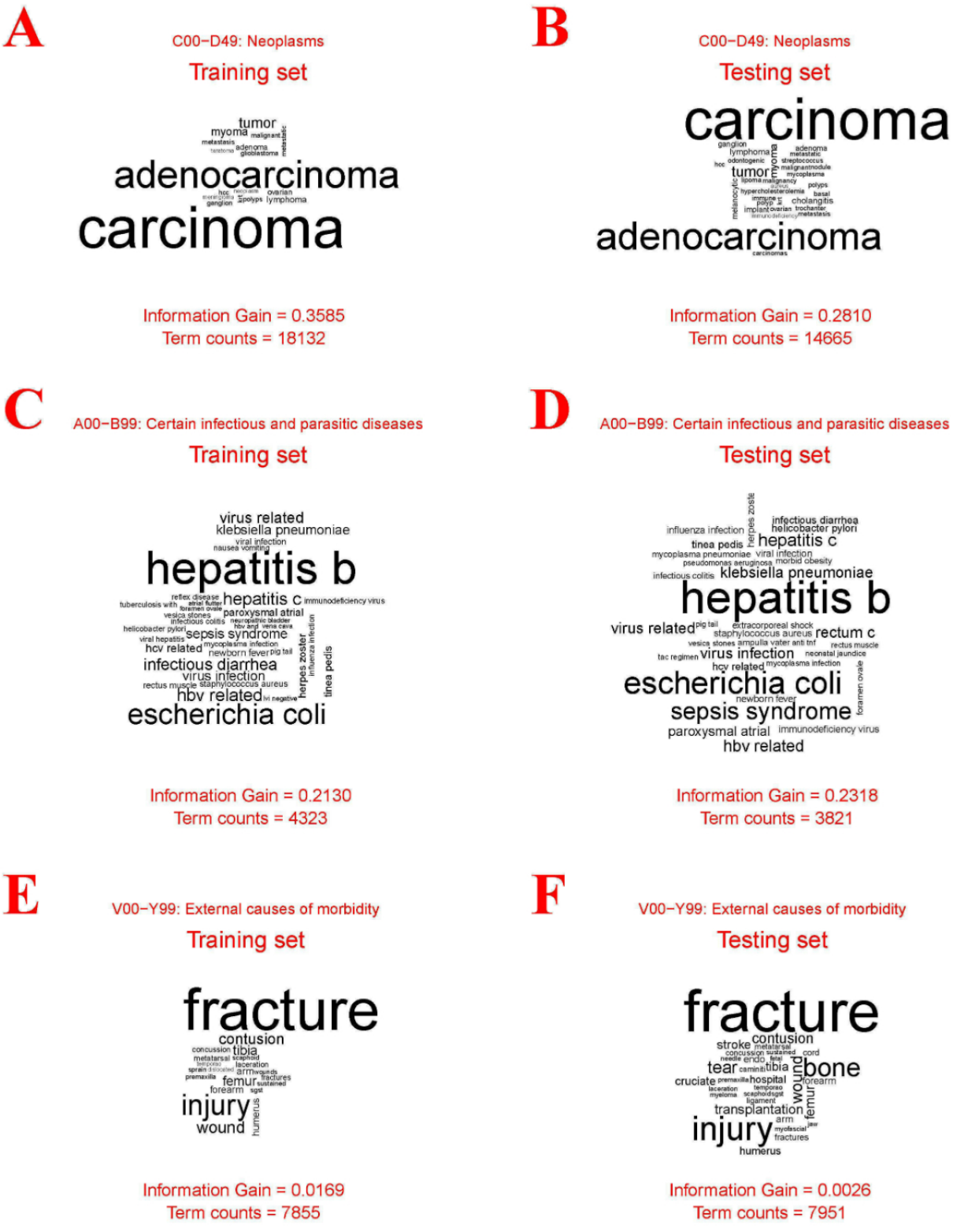


Figure 3. Information gains of the features extracted by the convolving filters in each classification task. AUC: area under the curve; IG: information gain.

Discussion

Principal Findings

The proposed method, which combines word embedding with a CNN, had a higher testing accuracy than all traditional NLP-based approaches, regardless of the situation. Further analysis showed that convolving filters had fuzzy matching abilities, which greatly reduced the data dimension for the final classification task. Moreover, the training AUCs of the traditional methods were very close to 1. This means that there was no possibility of improvement, and the larger difference between training set and testing set performances implies overfitting.

Arbitrary free-text medical narratives include many word combinations, and there is no good way of integrating similar terms using the current NLP pipelines. Previous studies have highlighted this issue and suggested that improvements are possible by dealing more effectively with the idiosyncrasies of the clinical sublanguage [7]. We believe that our proposal has an advantage in this respect. The used fuzzy matching technology offers a real chance of reducing the risk of overfitting. This is not surprising, as CNNs have achieved excellent results in some text mining tasks [13-15,22,27,28]. This study also demonstrated the advantages of using CNNs for free-text medical narrative classification.

Our proposed method not only increased the accuracy compared with traditional methods, but also can avoid troublesome data preprocessing. Our solution for avoiding troublesome data preprocessing is based on word embedding, which can learn semantics from external resources. The vocabularies are mapped to vectors of real numbers, and the word vectors for similar concepts are likewise close. In our work, a discharge note is converted into an n×50 matrix, where n is the number of words, and CNN classifies this matrix based on our designed convolving filters. Because the word vectors for similar concepts are likewise close in terms, convolutional layers effectively identified a large number of keywords in a convolving filter (data shown in Figure 2.). Finally, we used the document features extracted by these convolving filters to identify *ICD-10-CM* diagnosis codes. This simple idea effectively deals with the idiosyncrasies of the clinical sublanguage, so the proposed method does not require data preprocessing by external dictionaries.

All the classifiers used in this study performed poorly on V00-Y99 (external causes of morbidity) coding tasks, which may be attributed to sparse testing data (0.2%). A previous study found that classifier performance was better on common cancers than on rare cancers [2]. However, the performance of the proposed method was clearly better than that of traditional methods. The Q00-Q99 (congenital malformations, deformations, and chromosomal abnormalities) coding tasks were the next key point, as our method was obviously worse than traditional methods in these tasks. After further analysis, we found that the most common second-level *ICD-10-CM* diagnosis codes in Q00-Q99 are Q80-Q89 (other congenital malformations), and the words used in these discharge notes were really complex. This means that our CNN may have needed

more convolving filters to handle this issue. After we doubled the number of filters and retrained the CNN, the testing AUC greatly improved (testing AUCs of 0.9203 and 0.9235 in the cross-validation test and the real-world test, respectively). Hence, although a simple 1-layer CNN has already shown outstanding performance in our experiments, we believe that there are many opportunities to improve the performance of the proposed model.

All traditional term-based classifiers face the problem that emerging diseases cannot possibly be correctly classified. For example, influenza H1N1 could not possibly have been recorded in clinical narratives from 2000 to 2007, so term-based classifiers could not have been aware of the H1N1 pandemic of 2009 [3]. Our method can handle this problem using fuzzy matching technology. Although H1N1 was not recorded in discharge notes from 2000 to 2007, there was enough information to allow the machine to understand that H1N1 was an influenza subtype. In our pretrained GloVe model, H1N1 was very close to some influenza-related terms, such as "swine," "influenza," "flu," and "H5N1" (the cosine similarities were 0.835, 0.832, 0.831, and 0.716, respectively). Thus, we believe that convolving filters could still have correctly identified H1N1 and classified related discharge notes as A00-B99 (certain infectious and parasitic diseases), but more precise coding would have been difficult. Thus, retraining or incrementally updating the classifiers would still be necessary; otherwise, emerging diseases would be merged into similar disease categories. However, this is still an important breakthrough in the free-text medical writing classification task.

Previous studies described the classification methods used by human experts, and several rule-based approaches have demonstrated superior performance [3,33]. The only problem with rule-based approaches is that adding new diseases requires the development of new models and rules. RF models use an ensemble of decision trees, where each interior node is differentiated on the basis of 1 of the terms. We consider the similarity between RF and rule-based approaches to be higher than with the proposed CNN. The machine must imitate human behavior patterns to improve its correctness. The RF model showed better performance than traditional classifiers in most identification tasks (mean testing ranks of 3.000 and 3.190 in the cross-validation test and real-world test, respectively), possibly attributed to the RF model having a similar identification process to that of human experts. The proposed CNN architecture uses a logistic function for output, similar to a linear SVM, although nonlinear SVMs showed a lower training AUC, which may have been due to wrong assumptions about the relationship between features and the outcome. This evidence shows that the assumption of a linear relationship between extracted features and outcome is better than a nonlinear assumption, and the architecture of our CNN also follows this linear assumption in its last layer. However, rule-based approaches are more inclined to use positive terms than negative ones [3,33], so the architecture of RF or GBM is better than a linear classifier. The proposed CNN showed the highest accuracy; the key to success is not our network architecture but the fuzzy term matching technology. Fuzzy term matching reduces the hazard of overfitting, and the mean training AUCs for the RF and GBM models were higher than those for the other models, possibly indicating that overfitting is more risky in RF and GBM models. In summary, we consider that a deeper CNN may provide more accurate predictive ability. Further studies need to consider this to improve the performance of word embedding combined with a CNN.

Outbreaks of deliberate and natural infectious disease can lead to massive casualties unless public health actions are promptly instituted [34]. Thus, many countries have been building real-time infectious disease surveillance systems, such as the Real-time Outbreak and Disease Surveillance system [35]. The implementation principle of the Real-time Outbreak and Disease Surveillance system is through the structured *ICD* code, and it needs real-time manual identification by emergency physicians. However, this system cannot be extended to all diseases because a lot of resources are required. In addition to infectious diseases, other chronic diseases also need to be surveilled in real time [36]. Government health administrators need timely information to rapidly assess disease prevention and health protection priorities. A timely automated disease classification algorithm is required. Our proposed method provides a viable pipeline for implementing a disease surveillance system of all diseases. It not only improves classification performance but also avoids the inherent limitations of traditional methods. Subsequent studies can use this algorithm to further develop fully automated disease surveillance systems.

Limitations

Several potential limitations of this study should be acknowledged. First, we used only a 50-dimensional GloVe model to process our data, to reduce computing time. However, even a 50-dimensional model has better performance than traditional methods. Thus, we believe that this will not affect our result and that our proposal is a better solution for conducting free-text medical narrative coding tasks. Second, this study included discharge notes from only a single hospital, so we cannot confirm how well it would generalize to other data sources. Although this study only provided a feasibility assessment for extrapolation over time, we believe that it still demonstrated the superiority of our method. Third, this study conducted the classification task only in discharge notes. Discharge notes describe only the presence of the disease, but do not include negative statements. Our CNN architecture includes 3- to 5-gram phrase identifiers, but further studies are still needed to apply this approach to patient progress notes to prove its ability.

Conclusion

Our study showed that combining CNNs with word embedding is a viable analysis pipeline for disease classification from free-text medical narratives. Moreover, it showed outstanding performance compared with traditional NLP employing machine learning classifiers and may avoid troublesome data preprocessing. More complex CNNs could be used to further improve predictive performance, and future studies will not be limited by incomplete dictionaries. Because our data were collected from a single center, further studies can implement this algorithm in other hospitals. We hope our experiment will lead to a range of studies toward developing more efficient

automated classification approaches and that a large amount of unstructured information will be extracted from free-text medical writing. We have developed a Web app to demonstrate our work [37]. Public health surveillance systems would become more efficient, and government health administrators would be able to take timely and correct action for disease prevention and health protection. When previously unlabeled clinical records are labeled using such an automated approach, we can obtain more data-driven clues to help promote the progress of medicine. The health care field will then truly enter the age of big data.

Acknowledgments

This study was supported by the Smart Healthcare Project from the Medical Affairs Bureau Ministry of National Defense, Taiwan. Funding was supported by the Ministry of Science and Technology (105-2314-B-016-053) and Medical Affairs Bureau Ministry of National Defense (MAB-104-013). The authors appreciate the Medical Records Office at Tri-Service General Hospital for providing the unlinked data source.

Conflicts of Interest

None declared.

Multimedia Appendix 1

ICD-10-CM diagnosis code tutorial.

[PDF File (Adobe PDF File), 1MB - jmir_v19i11e380_app1.pdf]

Multimedia Appendix 2

Detailed training and testing AUCs and F-measures for the 5-fold cross-validation test.

[PDF File (Adobe PDF File), 54KB - jmir_v19i11e380_app2.pdf]

Multimedia Appendix 3

Detailed training and testing AUCs and F-measures for the real-world test.

[PDF File (Adobe PDF File), 54KB - jmir_v19i11e380_app3.pdf]

References

1. Lee LM, Thacker SB. Public health surveillance and knowing about health in the context of growing sources of health data. Am J Prev Med 2011 Dec;41(6):636-640. [doi: 10.1016/j.amepre.2011.08.015] [Medline: 22099242]
2. Koopman B, Zuccon G, Nguyen A, Bergheim A, Grayson N. Automatic ICD-10 classification of cancers from free-text death certificates. Int J Med Inform 2015 Nov;84(11):956-965. [doi: 10.1016/j.ijmedinf.2015.08.004] [Medline: 26323193]
3. Koopman B, Karimi S, Nguyen A, McGuire R, Muscatello D, Kemp M, et al. Automatic classification of diseases from free-text death certificates for real-time surveillance. BMC Med Inform Decis Mak 2015 Jul 15;15:53 [FREE Full text] [doi: 10.1186/s12911-015-0174-2] [Medline: 26174442]
4. Koopman B, Zuccon G, Wagholikar A, Chu K, O'Dwyer J, Nguyen A, et al. Automated reconciliation of radiology reports and discharge summaries. AMIA Annu Symp Proc 2015 Nov;2015:775-784 [FREE Full text] [Medline: 26958213]
5. Khachidze M, Tsintsadze M, Archuadze M. Natural language processing based instrument for classification of free text medical records. Biomed Res Int 2016 Sep;2016:8313454 [FREE Full text] [doi: 10.1155/2016/8313454] [Medline: 27668260]
6. Mujtaba G, Shuib L, Raj RG, Rajandram R, Shaikh K, Al-Garadi MA. Automatic ICD-10 multi-class classification of cause of death from plaintext autopsy reports through expert-driven feature selection. PLoS One 2017 Feb;12(2):e0170242 [FREE Full text] [doi: 10.1371/journal.pone.0170242] [Medline: 28166263]
7. Spasić I, Livsey J, Keane JA, Nenadić G. Text mining of cancer-related information: review of current status and future directions. Int J Med Inform 2014 Sep;83(9):605-623 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.009] [Medline: 25008281]
8. Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. J Mach Learn Res 2003 Feb;3:1137-1155 [FREE Full text]
9. Yih W, Toutanova K, Platt J, Meek C. Learning discriminative projections for text similarity measures. 2011 Jul Presented at: Fifteenth Conference on Computational Natural Language Learning; Jun 23-24, 2011; Stroudsburg, PA, USA URL: https://dl.acm.org/citation.cfm?id=2018965
10. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Dec Presented at: 2014 Conference on Advances in Neural information processing systems; Dec

9, 2013; Lake Tahoe, NV, USA URL: https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality
11. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. 2014 Oct Presented at: Conference on Empirical Methods in Natural Language Processing; Oct 26-28, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1162]
12. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc. IEEE 1998 Nov;86(11):2278-2324. [doi: 10.1109/5.726791]
13. Yih W, He X, Meek C. Semantic parsing for single-relation question answering. 2014 Jul Presented at: 52nd Annual Meeting of the Association for Computational Linguistics; Jun 23-25, 2014; Baltimore, MD, USA p. 23-25.
14. Shen Y, He X, Gao J, Deng L, Mesnil G. Learning semantic representations using convolutional neural networks for web search. 2014 Apr Presented at: 23rd International Conference on World Wide Web; Apr 7-11, 2014; Seoul, Korea p. 7-11. [doi: 10.1145/2567948.2577348]
15. Kim Y. Convolutional neural networks for sentence classification. 2014 Oct Presented at: Conference on Empirical Methods in Natural Language Processing; Oct 26-28, 2014; Doha, Qatar p. 26-28. [doi: 10.3115/v1/d14-1162]
16. Hornik K, Murdoch D. Watch your spelling!. The R Journal 2010 Sep;3(2):22-28 [FREE Full text]
17. Graff D, Cieri C. English Gigaword LDC2003T05. Philadelphia, PA: Liguistic Data Consortium; 2003. URL: https://catalog.ldc.upenn.edu/ldc2003t05 [accessed 2017-10-30] [WebCite Cache ID 6ubOmjzlO]
18. Hornik K, Buchta C, Zeileis A. Open-source machine learning: R meets Weka. Comput Stat 2008 May 14;24(2):225-232. [doi: 10.1007/s00180-008-0119-7]
19. Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. J Stat Softw 2008 Mar;25(5):125574. [doi: 10.18637/jss.v025.i05]
20. Butt L, Zuccon G, Nguyen A, Bergheim A, Grayson N. Classification of cancer-related death certificates using machine learning. Australas Med J 2013 Mar;6(5):292-299 [FREE Full text] [doi: 10.4066/AMJ.2013.1654] [Medline: 23745151]
21. Meyer D, Dimitriadou E, Hornik K, Leisch F, Weingessel A, Leisch F, et al. The Comprehensive R Archive Network. 2017 Feb. e1071: misc functions of the Department of Statistics, Probability Theory Group (formerly E1071), TU Wien URL: https://CRAN.R-project.org/package=e1071 [accessed 2017-10-30] [WebCite Cache ID 6uaY1A0It]
22. Al-garadi MA, Varathan KD, Ravana SD. Cybercrime detection in online communications: the experimental case of cyberbullying detection in the Twitter network. Comput Hum Behav 2016 May;63:433-443. [doi: 10.1016/j.chb.2016.05.051]
23. Candel A, Lanford J, LeDell E, Parmar V, Arora A. Deep learning with H2O, 3rd edition. Mountain View, CA: H2O.ai, Inc; 2015 Aug. URL: https://www.scribd.com/doc/312515027/Deep-Learning-with-H2O [accessed 2017-10-29] [WebCite Cache ID 6uaYPhzCU]
24. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013 Dec;7:21 [FREE Full text] [doi: 10.3389/fnbot.2013.00021] [Medline: 24409142]
25. Friedman J. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29(5):1189-1232.
26. Wolpert D, Macready W. No free lunch theorems for optimization. IEEE Trans Evol Comput 1997 Apr;1(1):67-82. [doi: 10.1109/4235.585893]
27. Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. 2014 Jun Presented at: 52nd Annual Meeting of the Association for Computational Linguistics; Jun 23-25, 2014; Baltimore, MD, USA p. 23-25. [doi: 10.3115/v1/P14-1062]
28. Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv 2015: 1510.03820. Ithaca, NY: Cornell University Library; 2015. URL: https://arxiv.org/abs/1510.03820v1 [accessed 2017-10-30] [WebCite Cache ID 6ubOIti8Q]
29. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014 Jun:1929-1958 [FREE Full text]
30. Chen T, Li M, Li Y, Lin M, Wang N, Wang M, et al. MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems. 2016 Presented at: Neural Information Processing Systems; Dec 10, 2016; Barcelona, Spain URL: https://arxiv.org/abs/1512.01274
31. Chai KEK, Anthony S, Coiera E, Magrabi F. Using statistical text classification to identify health information technology incidents. J Am Med Inform Assoc 2013 Sep;20(5):980-985 [FREE Full text] [doi: 10.1136/amiajnl-2012-001409] [Medline: 23666777]
32. Simpson A. Over-sampling in a deep neural network. arXiv 2015: 1502.03648. Ithaca, NY: Cornell University Library; 2015. URL: https://arxiv.org/abs/1502.03648 [accessed 2017-10-30] [WebCite Cache ID 6ubOQqAKk]
33. Muscatello DJ, Morton PM, Evans I, Gilmour R. Prospective surveillance of excess mortality due to influenza in New South Wales: feasibility and statistical approach. Commun Dis Intell Q Rep 2008 Dec;32(4):435-442 [FREE Full text] [Medline: 19374272]
34. Dembek ZF, Kortepeter MG, Pavlin JA. Discernment between deliberate and natural infectious disease outbreaks. Epidemiol Infect 2007 Apr;135(3):353-371. [doi: 10.1017/S0950268806007011] [Medline: 16893485]
35. Tsui F, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. J Am Med Inform Assoc 2003 Sep;10(5):399-408 [FREE Full text] [doi: 10.1197/jamia.M1345] [Medline: 12807803]

36. Mitra D, Hutchings K, Shaw A, Barber R, Sung L, Bernstein M, et al. Status Report – The Cancer in Young People in Canada surveillance system. Health Promot Chronic Dis Prev Can 2015 Jun;35(4):73-76 [FREE Full text] [Medline: 26083522]
37. Lin C. ICD-10 CM classificator. 2017. URL: https://chin-lin.shinyapps.io/icd10/ [accessed 2017-10-27] [WebCite Cache ID 6uX5IBFWP]

Abbreviations

AUC: area under the curve
CNN: convolutional neural network
GBM: gradient boosting machine
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
NLP: natural language processing
SVM: support vector machine
RF: random forest

Edited by G Eysenbach; submitted 20.07.17; peer-reviewed by S Tao, R Robinson, CM Chu; comments to author 07.09.17; revised version received 25.09.17; accepted 04.10.17; published 06.11.17.

Please cite as:
Lin C, Hsu CJ, Lou YS, Yeh SJ, Lee CC, Su SL, Chen HC
Artificial Intelligence Learning Semantics via External Resources for Classifying Diagnosis Codes in Discharge Notes
J Med Internet Res 2017;19(11):e380
URL: http://www.jmir.org/2017/11/e380/
doi:10.2196/jmir.8344
PMID:29109070

Original Paper

Public Awareness and Use of German Physician Ratings Websites: Cross-Sectional Survey of Four North German Cities

Stuart McLennan[1,2], MBHL, PhD; Daniel Strech[2], Dr med, Dr phil; Andrea Meyer[3], PhD; Hannes Kahrass[2], PhD

[1]Institute for Biomedical Ethics, Universität Basel, Basel, Switzerland

[2]Institute for History, Ethics and Philosophy of Medicine, Hannover Medical School, Hannover, Germany

[3]Division of Clinical Psychology and Epidemiology, Department of Psychology, Universität Basel, Basel, Switzerland

Corresponding Author:
Stuart McLennan, MBHL, PhD
Institute for Biomedical Ethics
Universität Basel
Bernoullistrasse 28
Basel, 4056
Switzerland
Phone: 41 612071786
Fax: 41 612071780
Email: s.mclennan@unibas.ch

Abstract

Background: Physician rating websites (PRWs) allow patients to rate, comment, and discuss physicians' quality. The ability of PRWs to influence patient decision making and health care quality is dependent, in part, on sufficient awareness and usage of PRWs. However, previous studies have found relatively low levels of awareness and usage of PRWs, which has raised concerns about the representativeness and validity of information on PRWs.

Objective: The objectives of this study were to examine (1) participants' awareness, use, and contribution of ratings on PRWs and how this compares with other rating websites; (2) factors that predict awareness, use, and contribution of ratings on PRWs; and (3) participants' attitudes toward PRWs in relation to selecting a physician.

Methods: A mailed cross-sectional survey was sent to a random sample (N=1542) from four North German cities (Nordhorn, Hildesheim, Bremen, and Hamburg) between April and July 2016. Survey questions explored respondents' awareness, use, and contribution of ratings on rating websites for service (physicians, hospitals, and hotels and restaurants) and products (media and technical) in general and the role of PRWs when searching for a new physician.

Results: A total of 280 completed surveys were returned (280/1542, 18.16% response rate), with the following findings: (1) Overall, 72.5% (200/276) of respondents were aware of PRWs. Of the respondents who were aware of PRWs, 43.6% (86/197) had used PRWs. Of the respondents who had used PRWs, 23% (19/83) had rated physicians at least once. Awareness, use, and contribution of ratings on PRWs were significantly lower in comparison with all other rating websites, except for hospital rating websites. (2) Except for the impact of responders' gender and marital status on the awareness of PRWs and responders' age on the use of PRWs, no other predictors had a relevant impact. (3) Whereas 31.8% (85/267) of the respondents reported that PRWs were a very important or somewhat important information source when searching for a new physician, respondents significantly more often reported that family, friends and colleagues (259/277, 93.5%), other physicians (219/274, 79.9%), and practice websites (108/266, 40.6%) were important information sources.

Conclusions: Whereas awareness of German PRWs appears to have substantially increased, the use of PRWs and contribution of ratings remains relatively low. Further research is needed to examine the reasons why only a few patients are rating physicians. However, given the information inequality between provider and consumer will always be higher for consumers using the services of physicians, it is possible that people will always rely more on interpersonal recommendations than impersonal public information before selecting a physician.

(J Med Internet Res 2017;19(11):e387) doi:10.2196/jmir.7581

KEYWORDS

physician rating websites; patient satisfaction

Introduction

When searching for a new physician, patients typically want to find a physician who is a *good* physician—clinically expert and, at the same time, interested in them, kind, courteous, empathetic, and caring [1]. Yet partly because of a lack of publicly available health care quality information, members of the public have traditionally had few ways of knowing who the *good* physicians are [2]. However, just as the public has used rating websites to find out information about the quality of other products and services, they are also increasingly using physician rating websites (PRWs) to obtain information about physicians, which allow the public to anonymously rate, comment, and discuss physicians' quality as a source of information for others [3-6]. In contrast to *top-down* public reporting approaches (Web 1.0), which only allow passive viewing of content (eg, the public reporting of health service performance using predetermined standards), PRWs represent a *bottom-up* public reporting approach (Web 2.0), which allows users to also generate content in the form of ratings and comments on physicians' quality as a form of electronic word of mouth [7-10].

Typically grounded in the assumptions of a theoretical consumer choice model [11], PRWs are a type of *public reporting activity* and have 2 key aims: (1) influencing patient decision making by increasing the chance that those patients who obtain information will choose better quality organizations or individuals [11-12] and (2) driving quality improvement by identifying aspects of care needing improvement so that changes can be made in practice [11-12]. A number of empirical studies indicate that PRWs are having some success in achieving these goals. For instance, a cross-sectional survey conducted in 2013 with a Web-based panel in Germany found that PRWs seem to have a meaningful influence (positive and negative) on choosing a physician [13]. A number of studies have also indicated that there is an association between PRWs and the quality of care. A cross-sectional survey of 2360 German physicians and other health care providers in 2015 reported that more than half of the responding providers used Web-based ratings to derive measures to improve patient care [14]. Furthermore, other studies have found a correlation between PRWs' ratings and objective measures of quality [15-18]. However, a number of shortcomings of PRWs have also been identified in the literature [19]. These include rating being anonymous, and therefore, not risk-adjusted and vulnerable to fraud, and a low number of ratings that are overwhelmingly positive [19]. This has raised concerns about the representativeness, validity, and usefulness of information on PRWs [19]. Indeed, without higher number of ratings, PRWs will continue to have limited value.

The utilization of the comparative quality information by health consumers depends on a range of factors, but at the most basic level, it requires consumers to first be aware of it [20], and it has been suggested that one reason for the low usage of PRWs might be that patients are still unaware of these websites [5]. However, 2 US studies published in 2014 by Hanauer et al found that whereas 74% of parents and 65% of adults in a nationally representative sample of the US population were aware of PRWs, only 28% of parents and 23% of adults had used PRWs [21,22]. Indeed, the level of PRW awareness reported by these 2 US studies were substantially higher than those found in previous German and English studies [9,13,23], however, the level of usage was comparable with the German studies [9,13]. This suggests that even if awareness of PRWs increases, there are more important factors behind the low level of PRW usage. For example, a Web-based survey of 1006 randomly selected German patients conducted in 2012 found that younger people, women, the highly educated, and people with chronic diseases were more likely to use PRWs [9]. However, regression analyses found that sociodemographic variables and health status alone did not predict PRW usage, but once psychographic variables and information-seeking behavior variables were added, it was found that higher education, poorer health status, higher digital literacy, lower importance of family and pharmacist for health-related information, higher trust in information on PRWs, and higher appraisal of usefulness of PRWs served as significant predictors for PRW usage [9].

With the most recent previous German studies regarding PRW awareness and usage having been conducted in 2012 and 2013 [9,13], there is a need to reexamine this issue to determine where future efforts to increase the level of PRW usage and ratings of physicians should be focused. Whereas both of the most recent German studies have drawn their samples from Web-based panels, this study will take a different approach to data collection and use a random sample of the general public, which should reveal a more generalizable view of the average population compared with panel data. This study aims to examine (1) the level of awareness and usage of PRWs among the general public and how this compares with other rating websites; (2) factors that predict awareness, use, and contribution of ratings on PRWs; and (3) attitudes toward PRWs in relation to selecting a physician.

Methods

This study was approved by Hannover Medical School's Research Ethics Committee on January 12, 2016. All participants signed an informed consent form.

Survey Implementation

A mailed survey was conducted between April and July 2016. A random sample was obtained from the Registry offices of four North German cities of various sizes, under paragraph 34 of the Federal Registry Act (Bundesmeldegesetz) that allows registration authorities to transfer data to other public bodies if certain criteria are met. Inclusion criteria for the random samples were that the person's place of residency was Nordhorn (53,285 residents, valid December 31, 2015), Hildesheim (101,667 residents, valid December 31, 2015), Bremen (556,326, valid December 1, 2015), or Hamburg (1,787,408 residents, valid December 31, 2015) and that the person was aged between 18 and 85 years. These cities were selected to enable participants from different sized cities to be recruited (small or rural area, medium city, large city, and extra-large city). To ensure participants' confidentiality and to allow for reminders to be sent to nonresponders, a unique identifier code was assigned to each participant from the random sample before data collection. The document with participants' identifying information and unique ID was accessible only to the study team, password

protected, and stored separately from data documents. Surveys with participants' unique IDs on them were mailed to a total of 1600 residents in Nordhorn (n=400), Hildesheim (n=400), Bremen (n=400), and Hamburg (n=400). Surveys were sent to residents of Nordhorn, Hildesheim, and Hamburg in the first week of April 2016, with a reminder sent to all nonrespondents 3 weeks later. Due to a late response from the Bremen Registry Office, surveys were sent to residents of Bremen in the first week of June 2016, with a reminder sent to all nonrespondents 3 weeks later. No incentives to participate in the study were provided. Fifty-eight surveys, which were returned because of out-of-date addresses or because participants had died or have a severe disability and are unable to read and write, were excluded from the study, leaving a total of 1542 surveys.

Survey Contents

Survey questions were primarily adapted from previous surveys conducted in Germany [9,12] and the United States [21,22]. The survey was pilot-tested with a convenience sample of 5 lay people using "think aloud cognitive interview" to ensure clarity and item comprehension [24]. Survey questions explored respondents' awareness, use, and contribution of ratings on rating websites for service (physicians, hospitals, and hotels and restaurants) and products (media and technical) in general and the role of PRWs when searching for a new physician. Questions concerning the importance of different information sources when searching for a physician, the usefulness of the information in PRWs, and how strongly this information influenced the decision regarding finding a physician used a 4-point Likert scale (eg, from "very important" to "not at all important"). Demographic questions asked for respondents' age, gender, marital status, education, whether they had previously been employed in health care, type of health insurance, whether they suffer from a chronic illness, and how often they had moved place of residence in the last 10 years (see Multimedia Appendix 1).

Data Analysis

For the questions concerning the awareness, use, and personal contribution of a rating on rating websites, there exists a cascade of questions where an inconsistent answer pattern could arise if a respondent answered "yes" to a question following a question they had answered "no" to. This inconsistency was solved by only including the first answer. Descriptive statistics included medians and means for continuous variables and percentages for categorical variables. Questions that used 4-point Likert response scales were dichotomized at the midpoint because sample sizes for some cells were often too small to be analyzed. Pearson chi-squared tests were used to analyze awareness, use, and contribution of ratings on rating sites, the role of PWRs when searching for a new physician, characteristics of respondents, and patterns of nonresponse. To test response rates between measures of the same subject (eg, awareness of PRW vs awareness of rating websites for hotels), we used the McNemar test. When comparing the percentages of participants using rating websites between two types of websites (PRWs and another), only those respondents who were aware of both types of websites were included. Similarly, when comparing the percentages of participants contributing a rating on two types of websites (PRWs and another), only those respondents who had used both types of websites were included. To assess potential predictors of the three outcomes—(1) awareness of PRWs, (2) use of PRWs, and (3) previously rated a physician—nine candidate predictors were preselected based on theoretical considerations and previous findings [9,13,21,22], including city, age, gender, marital status, education, previously worked in health care, health insurance, suffers a chronic illness, and number of times place of residence has changed in the last 10 years. We used three different models to test the impact of these predictors. First, we ran univariate logistic regression analyses for each predictor in a separate model, provided regression coefficients are unadjusted for all other predictors. Second, we used multiple logistic regression (MLR) models to test all predictors simultaneously, provided regression coefficients are adjusted for all other predictors in the model. MLRs, however, often suffer from overfitting, especially if the number of predictors is high relative to the number of cases and/or the frequency of the smaller group (in the case of dichotomous outcomes) [25], leading to models with low predictive accuracy when predicting new samples. To avoid such overfitting, we used a variable selection procedure, the least absolute shrinkage and selection operator (lasso), as a third model. In the lasso, model coefficients are deliberately shrunk by implying a penalty term to the binomial likelihood function when fitting the model. As a consequence, these models are somewhat more biased than those obtained from MLR but instead exhibit strongly increased predictive accuracy [26]. Thus, predictors whose coefficients from penalized regression have not been shrunk to 0 are likely to be predictive when replicating the study under consideration. Since the outcome variable was dichotomous, the lasso was based on a logistic regression model. The predictive accuracy of both the MLR and the lasso were determined by cross-validation [27]. We used Cohen kappa, computed between predicted and observed values, and the area under the receiver operating characteristics curves (ROC area) as measures of accuracy. A kappa value of 1 thereby denotes perfect predictive accuracy, whereas a value of 0 denotes random guessing. Accordingly, a value for ROC area of 1 denotes perfect predictive accuracy, whereas a value of 0.5 means random guessing. All predictor variables in the model were standardized before the analysis in any model, except for city and educational level.

Results

Characteristics of Respondents

Overall, a total of 280 completed surveys were returned, corresponding to an 18.16% (280/1542) response rate. Seventy-five of the completed surveys came from Nordhorn, 72 from Hildesheim, 62 from Bremen, and 71 from Hamburg. In addition, 169 formal refusals to participate in the study were received, of which 103 provided a reason for nonparticipation. Key reasons given for nonparticipation included no interest in the topic (n=21), no computer and/or Internet (n=10), not aware or do not use PRWs (n=9), health reasons (n=9), time reasons (n=5), and age reasons (n=5). Overall, 15.7% (44/280) of respondents were aged 30 years and less, 29.3% (82/280) were aged between 30 and 50 years, 38.6% (108/280) of respondents

were aged between 50 and 70 years, and 16.4% (46/280) were 70 years and older. Furthermore, 55.0% (154/280) of respondents were female, 58.9% (165/280) were married or in a civil partnership, 78.4% (218/278) had never been employed in health care, 81.4% (227/279) had public health insurance, 35.1% (98/279) suffered from a chronic illness, 29.7% (83/279) had changed their place of residence in the last 10 years 1 to 2 times, and 10.0% (28/279) had changed 3 or more times. Nonresponder analysis comparing all responders (n=280) with all nonresponders (n=1320) showed that gender composition did not significantly differ between responders and nonresponders (χ^2_1=1.9, *P*=.16), whereas mean age did (responders were older by 3.14 years on average, t_{1536}=2.68, *P*=.007); however, the effect size was small (*d*=0.18).

Awareness of Rating Websites

Overall, 72.5% (200/276) of the respondents were aware of PRWs (see Figure 1; Multimedia Appendix 2 presents the complete results). Respondents' awareness of PRWs was significantly lower than their awareness of rating websites for hotels and restaurants (χ^2_1=52.3, *P* ≤.001), technical products (χ^2_1=36.2, *P* ≤.001), and media (χ^2_1=18.8, *P* ≤.001), though significantly higher than that for hospitals (χ^2_1=33.9, *P* ≤.001). There was also a significant difference between participants' employment in the health care system and awareness of rating websites for physicians (χ^2_2=8.3, *P*=.02; 92% of the participants currently employed in health care were aware of PRWs, 59% of the participants previously employed in health care were aware of PRWs, and 72% of the participants never employed in health care were aware of PRWs).

Use of Rating Websites

Of the respondents who were aware of PRWs, 43.6% (86/197) had used PRWs (see Figure 2; Multimedia Appendix 2 presents the complete results). In comparison with other rating websites, this was significantly lower than the proportion who had used rating websites for hotels and restaurants (χ^2_1=44.2, *P*≤.001), technical products (χ^2_1=23.2, *P*≤.001), and media (χ^2_1=13.0, *P*≤.001), though significantly higher than for hospitals (χ^2_1=8.2, *P*=.004). There was also a significant difference between age groups and use of PRWs (χ^2_3=10.3, *P*=.02; 58%: 30 years or less, 54%: 30-50 years, 36%: 50-70 years, and 24%: 70 years and above).

Contributing to Rating Websites

Of the respondents who had used PRWs, 23% (19/83) had rated physicians at least once (see Figure 3; Multimedia Appendix 2 presents the complete results). This value was comparable with the proportion of ratings for the other websites except for hospitals where the respective value of 50% was significantly higher. Statistics comparing the proportions of ratings for physicians with each of the other rating sites were as follows: hotels and restaurants (χ^2_1=11.6, *P*≤.001), media (χ^2_1=4.8, *P*=.03), technical products (χ^2_1=3.5, *P*=.06), and hospitals (χ^2_1=0.2, *P*=.65).

Figure 1. Proportion of respondents who were aware of rating websites.

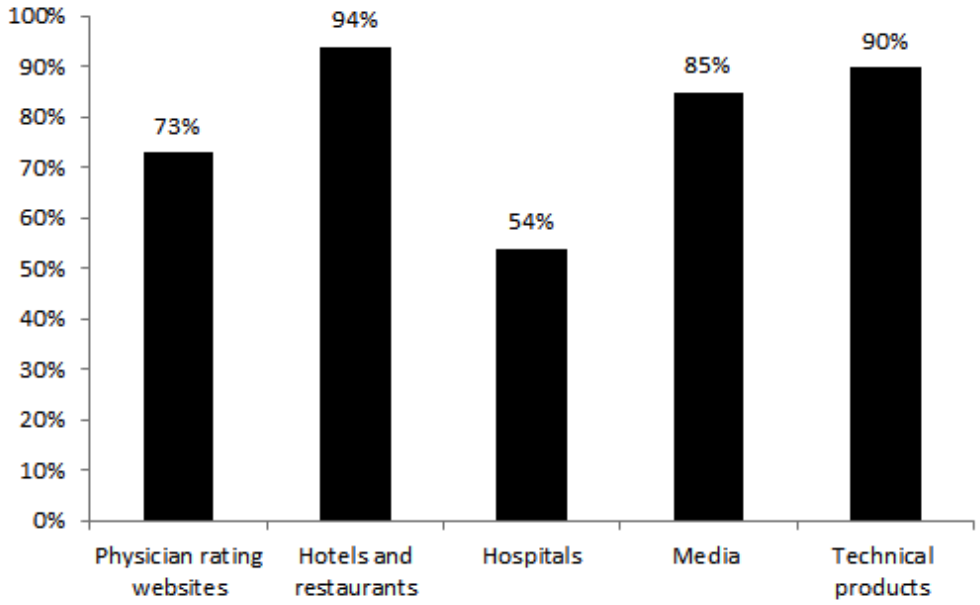


Figure 2. Proportion of respondents who had used rating websites.

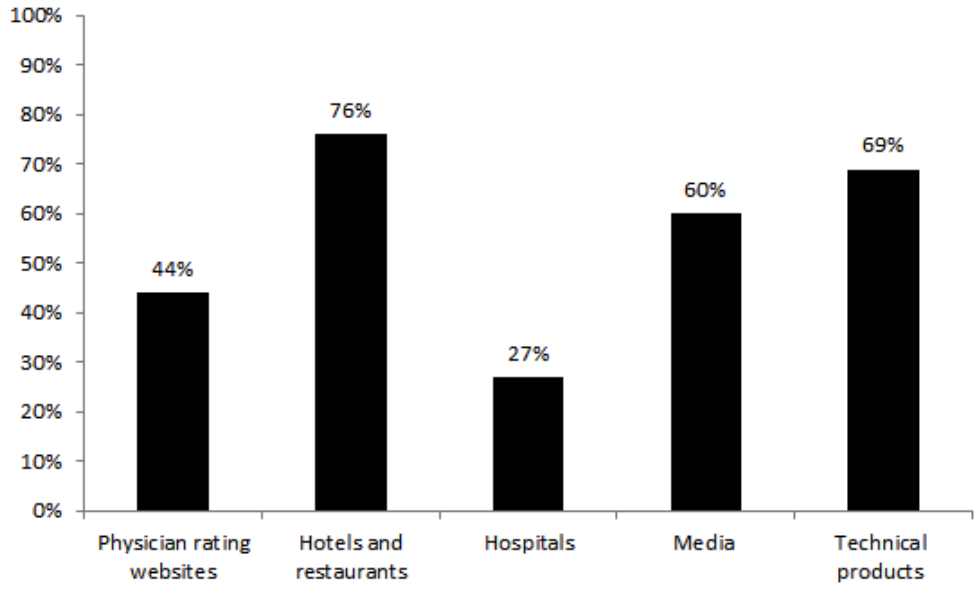
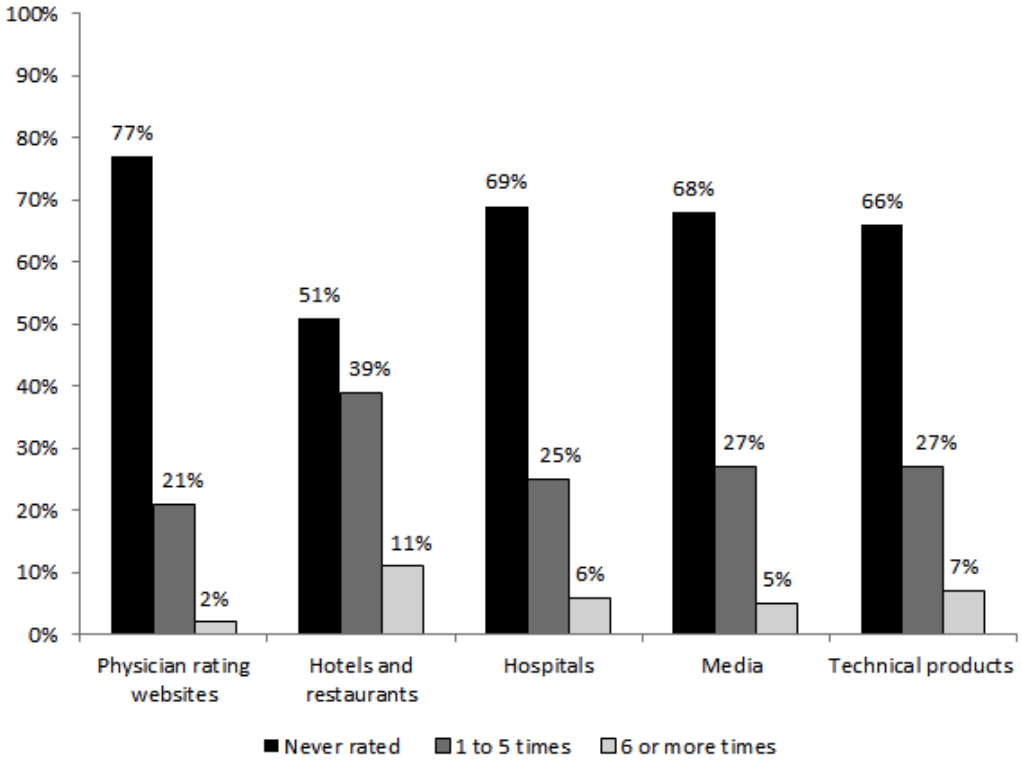


Figure 3. Proportion of respondents who had personally rated on rating websites.



Factors Predicting Awareness and Use of PRWs

Standardized coefficients of all predictors across all 3 outcomes were generally very low, suggesting that except for the impact of responders' gender (higher for females than for males) and marital status (higher if married or in a civil partnership than if not) on awareness of PRWs and responders' age on use of PRWs (higher for younger), no other predictors had a relevant impact. Complete results of the tests assessing potential predictors for the outcomes "already aware of PRWs," "previously used a PRW," and "previously rated a physician on a PRW" are shown in Multimedia Appendix 3. Results of cross-validation are shown in Multimedia Appendix 4. Those for "rated a physician on a PRW" were below 10%, whereas for the other two outcomes class imbalance was less severe. Predictive accuracy based on cross-validation was low for both the MLR and the lasso. Kappa values varied between 0 and 0.1 and ROC area values between 0.51 and 0.63 depending on the model and the outcome used. The corresponding values based on the sample data where the models were applied were somewhat higher, pointing to the degree of overfit in the MLR.

Table 1. Selecting a new physician.

Question	n (%)
When searching for a new physician, how important were the following information sources? (very or somewhat important)	
Family, friends, and colleagues	259 (93.5)
Other physicians	219 (79.9)
Practice websites	108 (40.6)
Physician rating sites	85 (31.8)
Business directories	51 (19.0)
Among those who answered "Yes" for used PRWs[a] : How useful did you find the information on physicians rating sites when searching for a physician?	
No experience	10 (12)
Very useful	9 (11)
Somewhat useful	28 (33)
Less useful	34 (40)
Not at all useful	4 (5)
Among those who answered "Yes" for used PRWs[a] : How strongly did the information on physician rating sites influence your decision regarding a new physician?	
No experience	9 (10)
Very strongly	7 (8)
Somewhat strongly	22 (26)
Less strongly	33 (38)
Not at all	15 (17)

[a]PRWs: physician rating websites.

Selecting a New Physician

Whereas 31.8% (85/267) of the respondents reported that PRWs were a very important or somewhat important information source when searching for a new physician, they significantly more often reported that other factors were very important or somewhat important information sources, with 93.5% (259/277) endorsing family, friends, and colleagues, 79.9% (219/274) other physicians, and 40.6% (108/266) practice websites (P ≤.001 for all three comparisons; see Table 1). There were significant differences between frequency of residency changes and importance placed on PRWs as an information source when searching for a new physician, with 24% considering PRWs as somewhat or very important among those who never moved, 41% among those who moved 1 to 2 times, and 46% among those who moved 3 or more times (χ^2_2=10.3, P=.006). For those respondents who had previously used PRWs, 44% (37/85) reported that they found the information on PRWs very useful or somewhat useful when searching for a new physician. Similarly, 34% (29/86) of the respondents who had previously used PRWs reported that the information on PRWs very strongly or somewhat strongly influenced their decision regarding a new physician, although this increased to 50% (11/22) among respondents who had previously rated a physician.

Discussion

Principal Findings

This study has resulted in three key findings. First, awareness of German PRWs was found to be high (72.5%), though the usage of PRWs (43.6%) and especially the contribution of ratings (23%) remain relatively low. Awareness, use, and contribution of ratings on PRWs were also significantly lower in comparison with all other rating websites, except for hospital rating websites. Second, respondents' age was the only relevant predictor of use of PRWs, with younger respondents more likely to use PRWs. Third, when selecting a new physician, the importance of factors such as family and friends as information sources were endorsed more frequently than PRWs.

Respondents' reported awareness of PRWs was substantially higher than the awareness of PRWs reported by previous German studies, which found 29% awareness in 2012 and 32% in 2013 [9,13]; however, it is similar to the 2 US studies published in 2014 by Hanauer et al, which found that 74% of parents and 65% of adults in a nationally representative sample of the US population were aware of PRWs [21,22]. It therefore appears that a lack of PRW awareness is no longer a key barrier to PRW usage (in Germany and the United States) [5]. The apparent increase in awareness of German PRWs may have been influenced by the low response rate, with responders more likely to already be aware of PRWs and the majority of nonresponders said they were not interested in the topic.

However, it is also potentially due to large German public health insurers launching their own PRWs. Germany's largest public health insurer Allgemeine Ortskrankenkasse (AOK) launched the PRW Arzt-Navigator in 2010, which was rolled out nationwide in May 2011. Two other large public health insurers, Techniker Krankenkasse (TK) and BARMER GEK, have also subsequently developed their own PRWs (TK-Ärzteführer and BARMER GEK-Arztnavi, respectively). These PRWs have received media attention that has likely led to a much greater public awareness of PRWs [28].

However, despite this increase in PRW awareness, the percentage of those who had used PRWs (among the responders who were aware of them) was substantially lower than previous German studies. For instance, whereas Emmert et al only reported that 25% of all respondents had used PRWs, their data show that among those who were aware of PRWs, 79% had used PRWs [13]. Similarly, while Terlutter et al only reported 26% of all respondents used PRWs at least once, their data show that among those who were aware of PRWs, 88% had used PRWs [9]. It is noteworthy that these numbers are even higher than the rate of users of the most common rating website in our survey (hotels and restaurants: 76%). The large difference between our study and the previous German studies likely reflects the increase in awareness of German PRWs and the different methodologies used in the 2 studies. When the previous German studies were conducted, most German PRWs were still in a growth phase and awareness of German PRWs was low; furthermore, both the Emmert et al and Terlutter et al studies used Web-based surveys and recruited participants from Web-based panels [9,13]. Our study, on the other hand, has been conducted when German PRWs are in a more mature phase, and awareness of German PRWs is much higher; additionally, our study used a mailed survey and a sample of the general public. Consequently, our study's findings more likely reflect the distribution between PRWs awareness and usage in the general public, although further research is needed to confirm this. Nevertheless, the 2 US studies by Hanauer et al involving the general public also found similar rates of PRW usage among those aware of PRWs: 38% and 37% [21,22].

Respondents' age was found to be the only relevant predictor of use of PRWs, with younger respondents more likely to use PRWs. This finding supports previous research [9,29] and the suggestion that the main impact of PRWs will likely come with the next "Facebook" generations, who use the Internet more and are more used to providing ratings on the Internet [30]. The use of PRWs is also likely influenced, in part, by the importance patients place on it as an information source when searching for a new physician. Whereas the vast majority of responders reported that family, friends, and colleagues were a very or somewhat important information source, only one-third of responders considered PRWs an important source. The 2 US studies by Hanauer et al reported that participants placed similar importance on family and friends as an information source (very or somewhat important: 85% and 89%, respectively) but substantially higher importance on PRWs (very or somewhat important: 59% and 62%, respectively) compared with the German participants in our study [21,22]. The large difference placed by US and German participants on the importance of PRWs as an information source when searching for a new physician may be partly because of differences in population mobility in the 2 countries; our study found that the proportion of German participants placing importance on PRWs increased with more changes of residency; however, research suggests that the US population is more mobile than European Union citizens [31]. Population mobility may be a fruitful area for future research to examine in relation to differences between countries in the use of PRWs.

Among the responders who had used PRWs, the percentage of those who had rated a physician was also much lower than previous German studies. For instance, whereas Emmert et al reported that only 11% of all respondents had rated a physician, their data show that among those who had used PRWs, 44% had rated a physician [13]. This higher rate of rating may also reflect the Web-based survey and Web-based panel used by Emmert et al. Previous international studies have found results similar to our study, with both US studies by Hanauer et al reporting that among those who had used PRWs, 23% had rated physicians [21,22]. As PRW users are not compensated for contributing ratings in any way, basic economic theory suggests that ratings were likely to be underprovided [32]. However, it is clear that users of other rating websites are contributing ratings more often. Whereas the lower awareness and usage of PRWs in comparison with other rating websites may remain a factor, there are clearly other factors that are leading patients not to rate their physicians given the low number of people who use PRWs going on to contribute their own rating. Of the responders who had used PRWs, under a third had contributed a rating. Two-thirds of PRW users are therefore effectively *free riders*, who are willing to use PRWs as a source of information but do not contribute ratings. However, without higher number of ratings, PRWs will continue to have limited value for every user. It is therefore not only a matter of reciprocity, but it is in the users' self-interest to contribute ratings. Many PRW users may not be aware of this, and PRWs may want to consider notifying users of the importance of contributing ratings.

However, it is also useful to consider the reasons why participants' awareness and usage of PRWs were all significantly lower in comparison with the other non–health care-related rating websites, which is consistent with the results of the 2 Hanauer et al studies [21,22]. In a recently published study by Rothenfluh et al, the choice-making processes of participants using the rating website TripAdvisor to select a hotel and the PRW Jameda to select a physician were explored [33]. Despite involving 2 service goods, major differences between the uses of the 2 rating websites were found. Whereas participants thought that choosing a physician was more important than choosing a hotel, participants spent less time searching for physicians than hotels and found choosing a physician much easier. Four themes were identified as being behind these differences. First, participants used a "trial and error" approach when selecting a physician, deciding only after a visit if they wanted to stay or switch to another, whereas they felt confident in making a definite choice regarding a hotel based on the information on the rating website. Second, participants expressed high trust in the medical profession and perceived that any of the listed physicians would be competent, while their

confidence in hotels was lower. Third, although participants felt confident in evaluating the quality of a hotel, they perceived that they had an inability to properly evaluate the skills and abilities of physicians. Finally, participants reported that interpersonal connection, gut feeling, and likeability played a huge role in selecting a physician, whereas the price and offered facilities were more important in relation to hotels [33]. Drawing from the economics literature, Rothenfluh et al distinguished between different types of services goods according to the level of information asymmetry between the provider and consumer and consequently suggested that these 2 service goods (hotels and physicians) cannot be treated equally because of their unequal attributes [33].

These findings suggest that the awareness and usage of PRWs may, in fact, always be lower in comparison with the other non-health care–related rating websites because the information inequality between provider and consumer will always be higher for consumers using the services of physicians (and hospitals) than it is for consumers using other services such as hotels and restaurants, and even more so for consumers buying products such as media and technical products. Whereas changes to the way in which PRWs are currently designed may make them more useful [33], it is possible that people will always rely more on interpersonal recommendations than impersonal public information before selecting a physician, and consequently not use PRWs as often as they would other rating websites. It is therefore perhaps not very surprising that participants endorsed far more frequently the importance of family and friends as information sources than PRWs when selecting a new physician, which is consistent with previous research [21,22].

Nevertheless, there is currently limited research examining the reasons why patients are not rating their physicians on PRWs, and more research is needed regarding this issue to identify barriers that may be addressed. A recently published study by Patel et al explored patients' views regarding rating general practitioners on PRWs, within the context of other feedback methods available in England [34]. Participants reported that they would not leave feedback on PRWs because of accessibility issues, privacy and security concerns, and because they felt feedback left on a website may be ignored [34]. Hanauer et al also asked participants in their 2012 US study to consider the implications of leaving negative comments about a physician [22]. Participants reported being concerned that their identity could be disclosed (34%) and that the physicians may take action against them for leaving negative comments (26%) [22].

Limitations

This study has a number of limitations that should be taken into account when interpreting the results. Responder bias may have influenced the results; however, as those who responded to our survey are likely to be generally more interested in the issue, the relatively low use of PRWs and low contribution of ratings should be taken seriously. Additionally, with a response rate of 18.1%, a generalization of the quantitative results to all the inhabitants of the four North German cities is likely not possible. Differences may also exist between other regions in Germany with respect to rating sites. However, the survey was mailed to a random sample of an average population from four North German cities of different sizes. The nonresponder analysis also found no significant difference in gender, and whereas responders were slightly older than nonresponders on average, the effect size was small. One of the reasons nonparticipants provided for not participating in the survey was that they were not aware of PRWs or do not use PRWs. If nonparticipants did not participate simply because they were not aware of PWRs, the average proportion of participants that is aware of PRWs is likely to be overestimated. However, out of the 169 refusals to participate in the study that were received, only 9 of them reported that they did not participate because they are not aware of PRWs or do not use PRWs. Furthermore, as there is a cascade of questions (awareness > use > contribution of ratings), there need not be an overestimation in relation to the use of PRWs and rating of a physician on a PRW at least once, since only participants who are actually aware of PWRs were included for these questions. We therefore do not think that this issue has significantly impacted our results. Responses were self-reported, and therefore we do not know the actual use of PRWs or what PRW participants visited. Additional research involving a larger sample would be desirable.

Conclusions

This study indicates that awareness of PRWs in Germany has substantially increased in recent years. This is a positive development and suggests that a lack of awareness is no longer a key barrier for PRW usage in Germany. Nevertheless, the level of usage of PRWs remains relatively low, and moving forward, the focus should be on a better understanding of the reasons why patients are not rating their physicians on PRWs, so that barriers that may be addressed can be identified. However, given the fact that the information inequality between provider and consumer will always be higher for consumers using the services of physicians, the awareness and usage of PRWs may, in fact, remain lower in comparison with the other non-health care–related rating websites. While changes to the way in which PRWs are currently designed may make them more useful, it is possible that people will always rely more on interpersonal recommendations than impersonal public information for selecting a physician.

Acknowledgments

The authors would like to thank Dr Sabine Bossert for her helpful suggestions regarding the development of the survey and Brita Sander for administrative assistance. The authors also thank the reviewers of the manuscript for their helpful comments and suggestions. This project was funded by the Hannover Medical School's Hochschulinterne Leistungsförderung (HiLF).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey.

[PDF File (Adobe PDF File), 57KB - jmir_v19i11e387_app1.pdf]

Multimedia Appendix 2

Awareness and use of rating websites.

[PDF File (Adobe PDF File), 27KB - jmir_v19i11e387_app2.pdf]

Multimedia Appendix 3

Factors predicting awareness and use of physicians rating websites.

[PDF File (Adobe PDF File), 30KB - jmir_v19i11e387_app3.pdf]

Multimedia Appendix 4

Model accuracy for multiple logistic regression model and the lasso based on 10-fold repeated cross-validation.

[PDF File (Adobe PDF File), 33KB - jmir_v19i11e387_app4.pdf]

References

1. Irvine DH. Everyone is entitled to a good doctor. Med J Aust 2007 Mar 05;186(5):256-261. [Medline: 17391090]
2. Paterson R. The Good Doctor: What Patients Want. Auckland, Newzealand: Auckland University Press; 2010.
3. 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]
4. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 06;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
5. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 01;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
6. Emmert M, Meier F, Heider AK, 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]
7. Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? J Interact Mark 2004;18(1):38-52. [doi: 10.1002/dir.10073]
8. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
9. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar 31;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
10. Lagu T, Hannon N, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
11. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality-of-care information? A systematic review. Med Care 2009 Jan;47(1):1-8. [doi: 10.1097/MLR.0b013e3181808bb5] [Medline: 19106724]
12. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]
13. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
14. Emmert M, Meszmer N, Sander U. Do health care providers use online patient ratings to improve the quality of care? Results from an online-based cross-sectional study. J Med Internet Res 2016 Sep 19;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]

15. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. Br Med J Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
16. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Web-based patient ratings and objective measures of hospital quality. Arch Intern Med 2012 Mar 12;172(5):435-436. [doi: 10.1001/archinternmed.2011.1675] [Medline: 22331980]
17. 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 Oct 17;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
18. Timian A, Rupcic S, Kachnowski S, Luisi P. Do patients "like" good care? Measuring hospital quality via Facebook. Am J Med Qual 2013;28(5):374-382. [doi: 10.1177/1062860612474839] [Medline: 23378059]
19. Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014 Feb 20;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
20. Hibbard JH, Berkman N, McCormack LA, Jael E. The impact of a CAHPS report on employee knowledge, beliefs, and decisions. Med Care Res Rev 2002 Mar;59(1):104-116. [doi: 10.1177/1077558702059001006] [Medline: 11877876]
21. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. 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]
22. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. J Am Med Assoc 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
23. 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. Br Med J Open 2012;2(6):e001493 [FREE Full text] [doi: 10.1136/bmjopen-2012-001493] [Medline: 23148340]
24. Willis GB, Artino Jr AR. What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. J Grad Med Educ 2013 Sep;5(3):353-356 [FREE Full text] [doi: 10.4300/JGME-D-13-00154.1] [Medline: 24404294]
25. Harrell Jr FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regressions, and Survival Analysis. New York: Springer; 2015.
26. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed). New York: Springer; 2009.
27. Kuhn M, Johnson K. Applied Predictive Modeling. New York: Springer; 2013.
28. Mihm A. Faz.: Frankfurter Allgemeine Ärzte bewerten Krankenkassen im Internet URL: http://www.faz.net/aktuell/wirtschaft/neues-online-portal-aerzte-bewerten-krankenkassen-im-internet-11906444.html [accessed 2017-11-03] [WebCite Cache ID 6ugz1cQK1]
29. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. BMC Health Serv Res 2015 Sep 26;15(1):416. [doi: 10.1186/s12913-015-1099-2] [Medline: 26410383]
30. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010 Dec 07;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]
31. Gáková Z, Dijkstra L. Ec.europa.eu.: European Union; 2008. Labour mobility between the regions of the EU-27 and a comparison with the USA URL: http://ec.europa.eu/regional_policy/sources/docgener/focus/2008_02_labour.pdf [accessed 2017-11-03] [WebCite Cache ID 6ugzMectH]
32. Fradkin A, Grewal E, Holtz D, Pearson M. Bias and reciprocity in online reviews: evidence from field experiments on Airbnb. 2015 Presented at: EC '15 Proceedings of the Sixteenth ACM Conference on Economics and Computation; June 15-19, 2015; Portland, Oregon, USA p. 641.
33. Rothenfluh F, Germeni E, Schulz PJ. Consumer decision-making based on review websites: are there differences between choosing a hotel and choosing a physician? J Med Internet Res 2016 Jun 16;18(6):e129 [FREE Full text] [doi: 10.2196/jmir.5580] [Medline: 27311623]
34. 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]

Abbreviations

AOK: Allgemeine Ortskrankenkasse
HiLF: Hochschulinterne Leistungsförderung
MLR: multiple logistic regression
PRWs: Physician rating websites
ROC area: area under the receiver operating characteristics curves

TK: Techniker Krankenkasse

Edited by G Eysenbach; submitted 24.02.17; peer-reviewed by D Hanauer, S Bidmon, U Sander, M Aebersold, T Kool, R Terlutter; comments to author 05.06.17; revised version received 23.08.17; accepted 11.10.17; published 09.11.17.

Please cite as:
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;19(11):e387
URL: http://www.jmir.org/2017/11/e387/
doi:10.2196/jmir.7581
PMID:29122739

Original Paper

Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies

Trisha Greenhalgh[1], FMedSci; Joseph Wherton[1], PhD; Chrysanthi Papoutsi[1], PhD; Jennifer Lynch[2], PhD; Gemma Hughes[1], MSc; Christine A'Court[1], FRCGP (UK); Susan Hinder[3], PhD; Nick Fahy[1], BA, BSc; Rob Procter[4], PhD; Sara Shaw[1], PhD

[1]Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]School of Health and Social Work, University of Hertfordshire, Hatfield, United Kingdom

[3]RAFT Research and Consulting Ltd, Clitheroe, Lancs, United Kingdom

[4]Department of Computer Science, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Trisha Greenhalgh, FMedSci
Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865 289294
Email: trish.greenhalgh@phc.ox.ac.uk

Abstract

Background: Many promising technological innovations in health and social care are characterized by nonadoption or abandonment by individuals or by failed attempts to scale up locally, spread distantly, or sustain the innovation long term at the organization or system level.

Objective: Our objective was to produce an evidence-based, theory-informed, and pragmatic framework to help predict and evaluate the success of a technology-supported health or social care program.

Methods: The study had 2 parallel components: (1) secondary research (hermeneutic systematic review) to identify key domains, and (2) empirical case studies of technology implementation to explore, test, and refine these domains. We studied 6 technology-supported programs—video outpatient consultations, global positioning system tracking for cognitive impairment, pendant alarm services, remote biomarker monitoring for heart failure, care organizing software, and integrated case management via data sharing—using longitudinal ethnography and action research for up to 3 years across more than 20 organizations. Data were collected at micro level (individual technology users), meso level (organizational processes and systems), and macro level (national policy and wider context). Analysis and synthesis was aided by sociotechnically informed theories of individual, organizational, and system change. The draft framework was shared with colleagues who were introducing or evaluating other technology-supported health or care programs and refined in response to feedback.

Results: The literature review identified 28 previous technology implementation frameworks, of which 14 had taken a dynamic systems approach (including 2 integrative reviews of previous work). Our empirical dataset consisted of over 400 hours of ethnographic observation, 165 semistructured interviews, and 200 documents. The final nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework included questions in 7 domains: the condition or illness, the technology, the value proposition, the adopter system (comprising professional staff, patient, and lay caregivers), the organization(s), the wider (institutional and societal) context, and the interaction and mutual adaptation between all these domains over time. Our empirical case studies raised a variety of challenges across all 7 domains, each classified as simple (straightforward, predictable, few components), complicated (multiple interacting components or issues), or complex (dynamic, unpredictable, not easily disaggregated into constituent components). Programs characterized by complicatedness proved difficult but not impossible to implement. Those characterized by complexity in multiple NASSS domains rarely, if ever, became mainstreamed. The framework showed promise when applied (both prospectively and retrospectively) to other programs.

Conclusions: Subject to further empirical testing, NASSS could be applied across a range of technological innovations in health and social care. It has several potential uses: (1) to inform the design of a new technology; (2) to identify technological solutions that (perhaps despite policy or industry enthusiasm) have a limited chance of achieving large-scale, sustained adoption; (3) to plan the implementation, scale-up, or rollout of a technology program; and (4) to explain and learn from program failures.

(J Med Internet Res 2017;19(11):e367) doi:10.2196/jmir.8775

KEYWORDS

diffusion of innovation; scale-up; program sustainability; implementation; complexity of innovations; business planning; NASSS framework; nonadoption, abandonment, scale-up, spread, sustainability framework; innovation adoption

Background

In 2004 and 2005, Greenhalgh et al published a multilevel framework for studying diffusion of innovations in health care, based on a cross-disciplinary systematic literature review [1,2]. A key finding was that most empirical studies had focused on short-term adoption of simple innovations by individual adopters. Studies of complex innovations (especially those requiring an organizational- or system-level adoption decision and a recurrent budget line); of the nonadoption and abandonment of innovations by individuals; and of local scale-up, distant spread, and long-term sustainability were sparse.

An update of that review in 2010 focused explicitly on organizational-level adoption and mainstreaming of technological innovations [3]. It identified some new literature on organizational-level routinization [4], but little new evidence on scale-up, spread, or sustainability—a finding that has been confirmed by other reviews since [5-9].

In recent years, technological innovation has moved apace and is now widely viewed as a significant potential contributor to health and wealth [10]. Yet the track record of technology programs, especially those that require major changes in organizations or the wider care system, is poor because of the combined problems of nonadoption and abandonment by individuals and difficulties with scale-up and spread [11]. While there is a growing general literature on the long-term sustainability of technology-supported change [12], studies of the sustainability of health and social care programs remain sparse.

These problems are illustrated by the paradox of telehealth (a term with contested definitions [13] but, broadly speaking, remote health care to the patient's home). Despite much policy-level talk of triggering a revolution in service delivery and many small-scale proof-of-concept examples, telehealth services are rarely mainstreamed or sustained [14]. Nonadoption and abandonment of telehealth technologies by their intended users is common [15-17]. A nationwide audit in Norway showed that, despite geographical remoteness, a history of early adoption of telehealth, a strong policy push, and adoption in principle by 75% of all hospitals, fewer than 1% of outpatient consultations in participating specialties were actually undertaken via telehealth in 2013 [18].

Poor uptake of technological innovations is often explained in terms of barriers and facilitators. In a recent review of telehealth in heart failure, for example, we identified technology barriers, patient barriers, staff barriers, team barriers, business and financial barriers, and governance and regulatory barriers [13]. This list (and a reciprocal list of facilitators) resonates with other barriers-and-facilitators studies in the literature, including electronic patient record systems [7,19], electronic prescribing [20,21] and surgical safety checklists [22]. Such studies are a useful start, but they fall short of *theorizing* the failure to adopt, scale up, spread, or sustain a technology-supported program.

As our 2004 review of diffusion of innovations found, it is not individual factors that make or break a technology implementation effort but the *dynamic interaction between them*. The more complex an innovation or the setting in which it is introduced, the less likely it is to be successfully adopted, scaled up, spread, and sustained [7,23,24]. These interactions are unlikely to be elucidated by the randomized controlled trial design that still dominates much health technology research [25]. Rather, we need studies that are interdisciplinary, nondeterministic, locally situated, and designed to examine the recursive relationship between human action and the wider organizational and system context [25].

Among others, Lupton [26], May and Finch [27], Nicolini [28], Pols and Willems [29], Maniatopoulos et al [30], and our own team [31] have used different approaches to produce rich theorizations of the unfolding fortunes of technology-supported programs in health care. But such academic outputs are not directly accessible to the clinician on the ward, the manager in the office, or the executive in the boardroom—nor, indeed, to the patient in his or her home. Other authors (whose work is summarized in the Results section below) have drawn on such literature to produce unifying frameworks aimed at informing the work of implementation, although no previous framework has focused explicitly on nonadoption, abandonment, scale-up, spread, or sustainability.

We aimed to produce an evidence-based, theory-informed, but also accessible and usable framework that would enable those seeking to design, develop, implement, scale up, spread, and sustain technology-supported health or social care programs to identify and help address the key challenges in different domains and the interactions between them.

Methods

Study Design

Figure 1 summarizes our design and methodology. The study had 2 parallel components: (1) secondary research (hermeneutic systematic review) to identify key domains and interactions, and (2) empirical case studies of technology implementation to

explore, test, and refine these, followed by a synthesis phase, and peer review and refinement of the draft framework.

Primary Research: 6 Empirical Case Studies

We selected a diverse sample of case studies from 2 research programs: Virtual Online Consultations: Advantages and Limitations (VOCAL) and Studies in Co-creating Assisted Living Solutions (SCALS), whose detailed methodology and ethical approval have been described elsewhere [25,32]. VOCAL (2015-2017, with an earlier set-up phase from 2011) was an in-depth study of the development, introduction, and local rollout of remote (video) consultations across 3 contrasting clinical specialties in a large, multisite UK hospital trust [32]. SCALS (2015-2020, with some data collected from 2013) is an action research study of the challenges faced by UK health and social care organizations who introduce technology-supported new service models; it includes examples from health care (eg, remote biomarker monitoring, video consultations, technologies for integrating care across organizations) and social care (safety alarms, global positioning system [GPS] tracking, care organizing apps) [25].

Case studies in VOCAL and SCALS involved qualitative interviews (with patients, clinicians, managers, technical designers, commercial partners, and—where relevant—investors), analysis of documents (correspondence, business plans, clinical records), ethnography (of technology use by patients or clients and staff, of meetings and events, and of technology design and functionality), and video recording of both ends of remote consultations [25,32]. Table 1 summarizes the subset of data from VOCAL and SCALS used for this study.

Figure 1. Study flowchart.

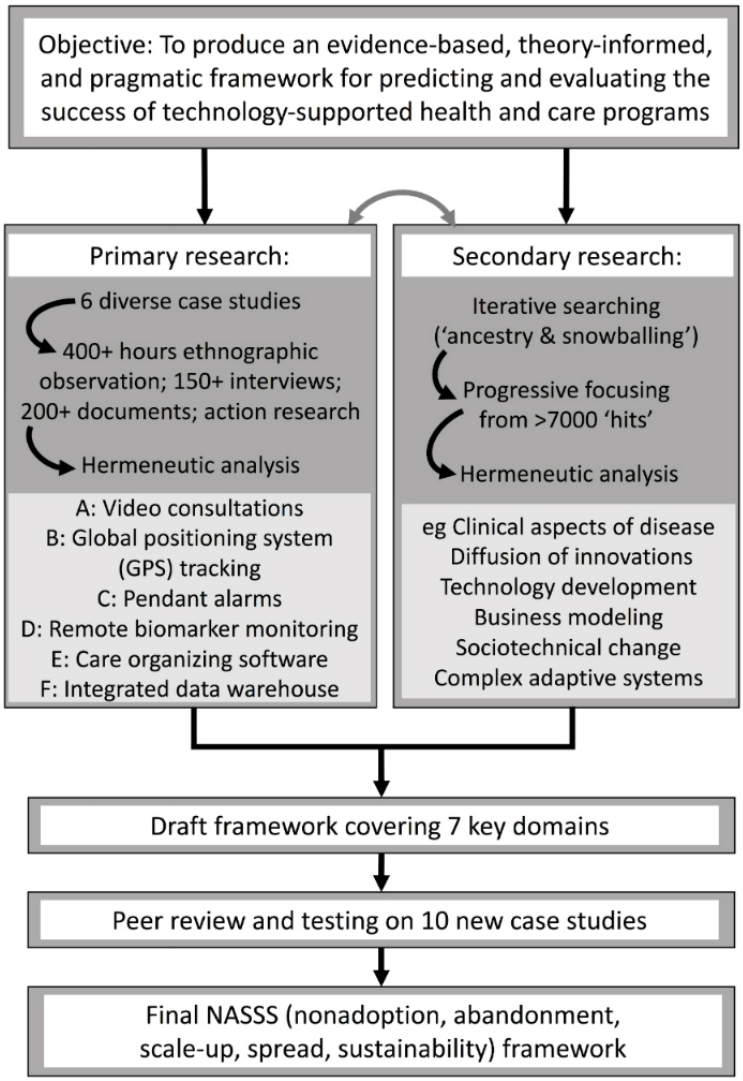


Table 1. Summary of data sources used in this analysis.

Study site or sites	Technology or technologies	Participants	Data sources
Video outpatient consultations			
1A. Acute hospital trust (3 specialties—diabetes, prenatal diabetes, cancer—on different sites) 1B. Nurse-led heart failure service run from community hospital	Skype (acute hospital) and FaceTime (community hospital)	1A. 24 staff (9 clinicians, 10 support staff, 5 managers); 30 patients 1B. 10 staff (8 nurses, 1 manager, 1 administrator); 8 patients Plus 48 national stakeholders and wider informants on remote consulting	35 formal semistructured interviews plus ~100 informal interviews; ≥150 hours of ethnographic observation; 40 videotaped remote consultations (12 diabetes, 6 prenatal diabetes, 12 cancer, 10 heart failure); ≥500 emails; 30 local documents such as business plans, protocols; 50 national-level documents
GPS[a] tracking for cognitive impairment			
2A. Social care organization in deprived borough in inner London, UK	GPS tracking devices supplied by 5 different technology companies, includes GPS tracking with virtual map and geofence alert functions	7 index cases; 8 lay caregivers; 5 formal caregivers; 3 social care staff; 3 health care staff; 3 call center staff	22 ethnographic visits and "go-along" interviews with index cases (~50 hours); 15 ethnographic visits with health and social care staff; 6 staff interviews; 5 team meetings; 3 local protocols
Pendant alarms			
3A. Health care commissioning organization in deprived borough in outer London, UK 3B. Social care organization in mixed borough in the Midlands, UK	In both sites, pendant alarms and base units were supplied by multiple different technology companies and supported by local councils, each with a different set of arrangements with providers and an "arms-length management organization" alarm support service	Site 3A. 8 index cases; 7 lay caregivers; 12 professional staff Site 3B. 11 index cases; 9 health and social care staff from frontline service delivery to senior board level; 3 representatives from telecare industry	50 semistructured and narrative interviews; 61 ethnographic visits (~80 hours of observation) including needs assessments and reviews; 20 hours of observation at team meetings
Remote biomarker monitoring in heart failure			
Acute hospital trusts in 6 different cities in United Kingdom	Tablet computer and commercially available sensing devices (blood pressure monitor, weighing scales, pulse oximeter)	7 research staff, including principal investigator and research coordinator for SUPPORT-HF[b] trial; 7 clinical staff involved in trial; 4 clinical staff not involved in trial; (to date) 18 patient participants and 1 spouse	1 patient focus group; 8 patient interviews; 24 additional semistructured interviews; SUPPORT-HF study protocol and ethics paperwork; material properties and functionality of biomarker database
Care organizing software			
5A. Health care commissioning organization in northern England 5B. National caregiver support charity in UK	5A. Web-based portal developed by small technology company for use by families to help them organize and coordinate the care of (typically) an older relative 5B. Smartphone app codesigned by caregiver support charity for same purpose	Product A: 2 technology developers and CEO[c] of technology company; 4 social care commissioners; 30 health and social care staff considering using the device; 4 users of the device, 1 nonuser. Product B (to date): 2 members of care charity (including CEO); 10 qualitative case studies of users undertaken by another academic team	22 semistructured and narrative interviews; 16 hours' ethnographic observations of meetings; autoethnographic testing of functionality and usability of devices; secondary analysis of third-party evaluation of Product B
Data warehouse for integrated case management			
1 acute hospital trust, 1 community health trust, 3 local councils, 3 health care commissioning organizations	Integrated data warehouse incorporating predictive risk modeling (in theory interoperable with record systems in participating organizations)	14 staff; 20 patient participants	14 semistructured interviews; 50 ethnographic visits (~80 hours); 12 hours' shadowing community staff; 4 hours' observation of interdisciplinary meetings; 12 local protocols or documents

[a]GPS: global positioning system.

[b]SUPPORT-HF: Seamless User-Centred Proactive Provision of Risk-Stratified Treatment for Heart Failure.

[c]CEO: chief executive officer.

Secondary Research: Hermeneutic Literature Review

Articles describing technology implementation frameworks and their applications were eligible if they (1) studied a technology that was perceived as new by intended users, (2) aimed (through the uptake and use of the technology) to improve service efficiency, or patient or client outcomes in health or social care; and (3) offered some kind of conceptual or theoretical framework. We were particularly interested in patient-facing technologies such as telehealth, but we also assessed other frameworks (eg, for health information systems) for transferable insights.

We began by selecting relevant studies from our hermeneutic literature review of telehealth in heart failure (covering 32 previous systematic reviews and 60 additional articles, including many that covered conditions beyond heart failure) [13]. We searched the reference lists of key studies [33-41]; we also put their titles into Google Scholar to identify 160 articles (surprisingly few) that had cited them subsequently, and manually screened these titles for relevance. We chose this "ancestry and snowballing" approach because initial database searching proved neither sensitive nor specific [42].

Having obtained few hits, we extended our search to the wider literature by tracking our original 2004 diffusion of innovations review [1]; we manually screened the titles and abstracts of over 4500 publications that had cited it. We did the same with 8 other highly cited reviews on the broader topic of innovation in health care [4-9,27,43] (around 3000 additional hits), using progressive focusing to limit the dataset. We favored authoritative reviews and added selected primary studies (characterized by strong theory, naturalistic methods, and rich detail, and including a focus on technology implementation). Where articles cited a specific theory, we obtained the original article describing that theory.

We used a simple data extraction form to summarize key aspects of each study (both theoretical and empirical). Using the hermeneutic (interpretive) methodology described in detail previously [13], we combined the findings of primary studies and previous reviews to generate a preliminary list of domains, potential interactions, and theoretical mechanisms.

Synthesis and Framework Development

All 6 case studies generated large amounts of qualitative and quantitative data, not all of which was relevant to the objective of this study. Our first task was to delineate a more focused dataset of individual index cases (patients or clients) along with relevant staff interviews, field notes, and background documents (Table 1). For each case study, we analyzed qualitative data thematically and produced an initial narrative summary of the case, which we refined in the light of emerging theoretical evidence from the literature review. We conducted data analysis of the empirical case studies in parallel with progress on the hermeneutic literature review; each influenced the other. Findings broadly coalesced around key domains of influence (the patient, the technology, staff, and so on), which informed the development of an initial framework and raised further questions about the implementation process.

Refinement, Peer Review, and Testing

We developed initial versions of the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) diagram and framework part way through the empirical work to guide our action research and inform cross-case theorization. As we applied the framework to real cases, challenges occurred that were not covered by it (eg, patients unable to use technologies because of comorbidities), so we searched more specifically for articles to inform additional domains and questions. We shared a near-final version of the framework with colleagues involved in 10 further large-scale technology-supported change programs (including email and video consultations in primary care; an online peer support network for people with mental health needs; remote biomarker monitoring in transplant patients; and an online tool for people to identify local services appropriate to their health and care needs); we further refined the framework in the light of their feedback.

Results

Table 1 shows the datasets for our empirical case studies. Below, we give a brief overview of the cases before summarizing our literature review and introducing the NASSS framework.

Empirical Case Studies

Case A: Video Outpatient Consultations

This case included 4 clinical services: 3 hospital-based services from VOCAL (young adult diabetes, prenatal diabetes, and cancer surgery, all using Skype; SS, unpublished data, 2017) and 1 community-based from SCALS (a nurse-led heart failure service run from 4 community hospitals, using predominantly FaceTime). In each, patients judged "appropriate" for video consultations by the doctor or nurse were offered this option. National policy makers viewed video consulting as a way of delivering health care efficiently to an aging population with rising rates of chronic illness. But the reality of establishing such services in busy and financially stretched public sector organizations proved far more complex and difficult than anticipated; progress was slow and required multiple organizational workarounds. Technical challenges in setting up video consultations with patients were typically mundane but potentially prohibitive (eg, forgotten passwords, poor connectivity, outdated software). When clinical, technical, and practical preconditions were met, video consultations appeared safe and were popular with both patients and staff, although only some clinicians agreed to participate.

By the end of the study period, video consultations had been abandoned in the prenatal diabetes service and put on hold in the community heart failure service, but the young adult diabetes and cancer surgery services were conducting around 20% of follow-up consultations remotely. In the (extremely busy) prenatal diabetes clinic, video consultations aligned poorly with a context involving multidisciplinary teams (patients were typically seeing multiple clinicians across departments) for a relatively short-term but high-risk condition and in the absence of integrated records (paper medical notes were held by the patient so not physically at hand for the clinician). In the heart failure clinic, the physical examination (eg, heart rhythm, leg

edema) that the nurses considered essential was not easy in the remote environment (although sometimes possible with patient and caregiver assistance); multimorbidity and polypharmacy were common and in most cases the perceived risks and uncertainties associated with remote consulting were considered to outweigh the benefits.

Case B: GPS Tracking in Cognitive Impairment

Electronic tracking through GPS is used to monitor people with cognitive impairment who "wander" outside the home. We worked with a public sector social care organization to implement and adapt GPS tracking devices and a linked monitoring service for such individuals (of whom 11 were considered eligible and 7 assented). In what were typically very complex care contexts, GPS devices were useful to the extent that they aligned with a wider sociotechnical care network that included lay caregivers, call centers, and health and social care professionals. In this context, "safe" wandering was a collaborative accomplishment that depended on the technology's material features, affordances, and aesthetic properties; a distributed knowledge of the individual and the places they wandered through; and a collective and dynamic interpretation of risk. Each index case required a high degree of tinkering (including customization of the device, liaison with the technology supplier, and adjustment of work routines) to achieve a solution that was acceptable. Despite this, only 3 individuals were still using the technology by the end of the 18-month study period.

Case C: Pendant Alarms

Pendant alarms (worn around the neck or on a wrist strap and connected to a remote call center) were the only patient-facing technology in widespread use in our dataset. Both study sites had a well-established sociotechnical infrastructure that included a named care team with expertise and local knowledge. Supply of a pendant alarm was typically initiated by a public sector organization and involved a local technology supplier to fit it, with or without support from an age charity. Clients could also self-refer. The setup usually depended on a network of lay caregivers available to respond to a summons; a safe box was usually installed containing a key so the rescuer could let themselves in, or an emergency response (eg, ambulance or 24/7 social care) was summoned as required. Users paid a set-up fee (around £25; US $40) plus a small weekly support fee (around £4.50; US $7), although some local care providers offered this service free of charge (eg, to people in receipt of welfare benefits). In almost all cases, the individual had multiple and complex needs (physical, cognitive, social) and was using multiple technologies in addition to the alarm [44].

In many but not all cases, activation of the pendant alarm led to help arriving promptly. On some occasions, there was a mismatch: the alarm was triggered when there was no objective need (perhaps by accident) or, more commonly, not triggered when caregivers felt it should have been—because the individual did not want to trouble anyone, did not believe the problem was serious or urgent or was unable to activate the device (eg, during a fit), or was not wearing the alarm at the time of the crisis. Sometimes, call center operators made judgments and put in "emotional work" to support the caller without alerting their relative or reassured them while help was on its way [45]. In 1 site, use of pendant alarms had evolved in that some lonely people (especially those with cognitive impairment) were being *encouraged* to press the alert button and talk to a call center operator even when there was no emergency, in order to reduce call-outs of the emergency services.

Case D: Remote Biomarker Monitoring (Telehealth)

This case study involved cardiology departments in 6 UK hospitals, each implementing biomarker monitoring (weight, blood pressure, heart rate) for heart failure as part of a multicenter randomized controlled trial (Seamless User-Centred Proactive Provision of Risk-Stratified Treatment for Heart Failure [SUPPORT-HF]). The tablet technology supplied to patient participants had been developed using a codesign methodology [46]. Participants in both arms of the trial received the technology and automated feedback messages (eg, if results went outside preset parameters); in the intervention arm, the patient's family physician was alerted to out-of-range results and offered suggestions for changes in therapy, whereas in the control arm, results were made available on a Web portal for the patient's physician to access if they chose to. Staff at the different SUPPORT-HF sites engaged variably with the study, sometimes leading to slower than predicted recruitment. A minority of clinicians were reluctant to refer patients or engage with the trial protocol, citing "previous bad experiences with telehealth," concern that a remote monitoring service would threaten their jobs, or a perception that patients "deserved better." Patient participants expressed a range of views about remote biomarker monitoring; some took an active interest in their readings, engaged enthusiastically with the feedback they received, and found this monitoring reassuring. Others found the experience confusing and did not know (or wish to know) what the numbers meant. In some cases, a research nurse known to the participants provided (unofficial) telephone support to maintain engagement. Another problem with the remote monitoring service was the variability of broadband speed outside the main cities, which meant that more than half of potentially eligible participants in 1 site could not be included in the study.

Case E: Care Organizing Software

This case study followed the very different fortunes of 2 software products, each designed to help relatives and friends (and sometimes professional staff as well) organize tasks and visits for someone with health or care needs. Product A, a Web portal, had been developed in-house by a small software company, based on a previous caring experience by one of the company staff. The business model was to sell the product to care organizations who would then provide it to their clients for free. The developer did not initially anticipate that either intended end users or participating care organizations would need any training or ongoing support to use the portal. Product A was not successful during the study period; fewer than 5 families were ever identified as actively using it.

Product B was a smartphone app (with linked Web portal) that had been developed via publicly funded research and development using codesign methodology by a national caregivers' charity. The charity had previously identified a need

for such software; they worked with a specialist app developer company and carefully selected pilot families. From the outset, the charity recognized that caregivers would need to be made aware of the product through mass mailing and to be actively invited and supported to use it, and that a helpdesk service would be needed. The app was made available commercially (via the App Store) for £2.99 (about US $5). Users signed up gradually but steadily; there was no tipping point, but at the time of writing over 1000 families are using the product through the care charity; in a preliminary evaluation, most spoke highly of it (and of the charity support).

Case F: Integrated Case Management Via Data Sharing

Case management is a way of organizing health and social care services through assessment and care planning by multidisciplinary teams with the aim of managing the growing challenge of emergency hospital admissions (and readmissions) in older people with multiple health and care needs. To avoid the high human and financial cost of such admissions, coordinated action and frequent dialogue between primary care providers, secondary care providers, and social care and other formal and informal caregivers is often needed. In the SCALS study, 1 site had introduced an integrated data warehouse incorporating a predictive risk modeling tool to automate the identification of people at high risk of hospital admissions through risk stratification, and to facilitate shared access to care plans in efforts to achieve integrated care. Although the data warehouse was part of business as usual in this site, in practice people at high risk of hospital admission were identified through a combination of risk stratification and clinical judgment. Care plans were shared in a range of ways, sometimes through the integrated data warehouse and sometimes bypassing it.

Literature Review

Our search for evidence-based approaches to guide our empirical work on the above case studies identified 28 technology implementation frameworks, informed by several theoretical perspectives, which we sorted into a simple taxonomy [13,43,47-72]; see Multimedia Appendix 1.

A key limitation of many previous frameworks was the lack of detailed analysis of the condition or problem for which a new technology was [part of] the intended solution. Some assumed a "textbook" condition—simple, isolated, easily characterized, and amenable to management by algorithm or protocol using a one-size-fits-all (or minimally customizable) technology. Yet there is much empirical evidence that the health and care needs of real people are extremely heterogeneous, even when they have the "same" condition. For example, Tait et al's case-by-case analysis of patients in a heart failure clinic found that every one of them required significant customization and ongoing adaptation of the care package recommended in the guideline [73].

A prominent policy prediction, typically couched in the language of "empowerment," is that remote technology will make care more efficient by encouraging self-management of chronic conditions [74]. But as May et al have pointed out in their burden of treatment theory, shifting the work of care from clinic to community places new demands on the sick (and hence raises ethical questions) [75]. Depending on the condition, such work may also be physically or cognitively impossible. The anchored, realistic, cocreative, human, integrated, and evaluated (ARCHIE) framework derived from our earlier empirical work on assisted living technologies emphasized the diverse manifestations of multimorbidity and social care need; it recommends commencing with a realistic assessment of the nature (and likely progression) of the condition and a focus on what matters to the user [47].

No previous framework explicitly considered inequalities in access, uptake, and use of health and care technologies by age, sex, socioeconomic status, or ethnic group, although previous empirical studies have highlighted substantial differences across such groups [76]. Chronic health conditions and care needs are strongly patterned by social determinants. For example, type 2 diabetes, heart failure, depression, cognitive impairment, and general frailty are all more than twice as common in the poorest and least well-educated quintiles of society as in the richest and best educated [77]. The poor may also have less rich social networks, less reliable access to broadband, lower digital literacy, and greater likelihood of having problems such as debt or unsuitable housing [77,78].

Most previous frameworks addressed the material properties of the technology, such as its physical features, functionality, and interoperability. Few considered its symbolic properties (some technologies—such as mobile phones—have connotations of youth, progress, and friendship; others—such as GPS tracking devices or pendant alarms—symbolize dependence or external control). One or two frameworks considered what knowledge or skills (and hence training and support) were needed for intended users to be confident in operating the technology.

A question addressed tangentially or not at all by previous frameworks was *what kind of knowledge* does the technology generate? For example, telehealth technologies fall into two broad categories. Remote monitoring devices transmit objective biomarkers such as weight, blood pressure, and oxygen saturation, and responses to closed questions on symptoms and compliance, and perhaps also transmit instructions or educational messages (what Pols has called "cold" telehealth [79]). Remote communication devices create the possibility for more conventional conversations between patients and clinicians by telephone or video ("warm" telehealth). These different technologies bring very different kinds of knowledge into play and, by design or default, exclude other knowledge and influences from the frame.

Very few frameworks in our sample included an assessment of whether a technology was likely to be *worth* introducing—that is, its value proposition. "Value" means different things to different stakeholders; it has parallels to Rogers' term "relative advantage' (the extent to which a potential adopter believes that the innovation is better than what has gone before [80]). From the patient's perspective, there is often a trade-off between the potential benefits of technologies, their costs (and the person's willingness and ability to contribute to these), the work required to use them (and the person's capacity to do so), and the desirability of medicalization and surveillance [81].

Lehoux et al distinguish between a health technology's upstream value as viewed by investors (especially the business case for generating profits, further spin-offs, and highly qualified jobs), drug and device regulators (preliminary evidence of efficacy and safety), and financial regulators (auditable business processes and governance), and its downstream value as viewed by clinicians and policy makers (including its impact on patients and health care costs) [82,83]. Health technology development is often characterized by poor alignment between supply-side and demand-side value [82,84,85].

In previous frameworks, technology adoption by health care staff was most commonly theorized using Davis's technology acceptance model, comprising perceived usefulness, perceived ease of use, and attitude toward the technology [86]; or Bandura's social learning theory, the relevant aspect of which is that people learn by observing and imitating the behavior of others [87,88]. Critics of the technology acceptance model have argued that it fails to account for human and social change processes [89]. Sociological theories of technology adoption, which emphasize the norms and expectations associated with different social positions and professional groups [26,27,31], were not extensively used in previous frameworks, with the exception of May and Finch's normalization process theory [27], of which relational integration—how the technology affects human relations such as the doctor-patient relationship—is one component.

In considering our own case studies, we were drawn to sociological theories because the new technologies often had implications for staff identity, professional commitments, and scope of practice. Acceptance by professional staff may be the single most important determinant of whether a new technology-supported service succeeds or fails at a local level [17,36,39,79,90,91]. Local champions appear key to persuading their peers that the technology-supported service is effective, safe, and "normal" (ie, professionally appropriate) [17,92]. We have previously developed a theoretical model of clinician resistance to new health care technology made up of 4 elements: resistance to the *policy* reflected in the technology (eg, a policy of shifting the work of disease management from professional to patient); resistance to the *sociomaterial constraints* (eg, clunkiness, dependability) of the new technology; resistance to *compromised professional practice* (eg, less scope for exercising judgment); and resistance to *compromised professional relationships* (eg, a perception that a remote interaction is less professional than a face-to-face one) [93].

One framework in our sample addressed technology acceptance by patients. The digital health engagement model was based on both burden of treatment theory and normalization process theory [48]. It proposes 4 key influences on whether an individual will engage with a health technology: personal agency and motivation (which is affected by aspects of their illness); personal life and values; the engagement and recruitment approach taken by those seeking to promote the technology; and the quality of the health technology.

Two frameworks drew on DeLone and McLean's classic theoretical model of information system success [94]. This considers system quality, information quality, usage attributes, user satisfaction, individual impact, and organizational impact [49,50]. While this framework had some resonance with our empirical data, it did not address the patient-facing aspects of health and care technologies, nor did it address contextual influences or change over time (we classified it as a static framework), and hence did not help us with our study of scale-up, spread, and sustainability.

Surprisingly few frameworks considered the organizational setting. Some antecedent characteristics of organizations have been shown to support innovation at an organizational level [1]. These include a devolved organizational structure (with each department or unit able to make semiautonomous decisions), significant organizational slack (that is, spare resources that can be channeled into new projects), and strong leadership, good managerial relations, a risk-taking climate (staff are rewarded rather than punished for trying things out), opportunities for sense making (that is, collectively arguing out the meaning of an innovation [95]), and what is known as absorptive capacity: "a set of organizational routines and processes by which [organizations] acquire, assimilate, transform, and exploit knowledge to create a dynamic organizational capacity" [96]. These organizational determinants of innovation align with complexity theory's emphasis on local adaptation and the need for creativity to address unique emerging issues [97]. A specific innovation is more likely to be taken up if there is strong tension for change, good innovation-system fit (that is, the innovation fits well with existing work and routines), widespread support for (and limited opposition to) the innovation, and systematic assessment of the implications [1].

Another important aspect of implementation omitted by most previous frameworks was the health or care organization's business model for introducing the technology. This includes the resources to support the model, key partners and relationships, the transaction mechanism (how will the organization interact with the supplier?), the value structure (how and when will value, including benefits for patients, be created and investment costs recouped?), and organizational design issues (what changes in organizational structure and processes are required or presumed by the new technology?) [11,14]. A review by van Limburg et al highlighted the financial and business challenges associated with eHealth technologies, including their (typically) fragmented deployment, multiple stakeholders and interdependencies, lack of recognition of the ongoing work of implementation, and an overreliance on the results of experimental efficacy trials [11].

Most, but not all, previous frameworks considered how a new technology would fit with existing organizational routines (defined as recurrent, collective patterns of interaction that both coordinate and control organizational work [4]). Technologies create opportunities for developing new routines and care pathways, but they also disrupt existing patterns of team interaction in ways that can prove more complex than initially anticipated [98]. There is, almost inevitably, a crucial gap between the nuanced, flexible, and often unpredictable nature of human activity and what it is possible to deliver technically. This is especially crucial when considering something as complex and exception filled as clinical work (Case A) [99]. As Grudin (cited by Symon et al [100]) put it:

> *Work processes can be described in two ways: the way things are supposed to work and the way they do work. Software that is designed to support standard procedures can be too brittle.* [page 25]

Cherns' classic theory of sociotechnical design, originally developed in the 1970s, is built on the principle that introducing technologies in an organization is a social process that depends on values, mindsets, and engagement, as well as on clear and extensive communication about what changes are occurring and why [101]. It is also an evolutionary process (sociotechnical systems are grown, not built), hence best achieved by early and active input of frontline workers into the [re]design of work routines—a principle that has long been recognized (but rarely adequately applied) in health care [100,102]. Also highly relevant to the health and care environment is what Weick called "technology as equivoque" [103] and Orlikowski (drawing on earlier work on the social construction of technology) called "interpretive flexibility" [104]: a technology introduced into an organization is open to multiple interpretations; successful embedding will require opening up a space for dialogue, listening to concerns, and allowing people time to argue out the challenges and learn from the experiences of others before a "closure" over possible interpretations is reached [88,105]. As Stewart and Williams [105] stated,

> *Innovation is not restricted to the prior design of an artefact, but continues as artefacts are implemented and used (innofusion). Supplier offerings are inevitably incomplete in relation to the complex, heterogeneous and evolving requirements of users; work needs to be done by specific users to incorporate these generic solutions to their particular contexts and practices (domestication).* [page 195]

Surprisingly, few previous frameworks in the health and social care literature have attempted to capture this insight, which partly explains why on-the-job training in technology use and ongoing helpdesk support are key to the implementation process [1].

A prominent theme in our empirical findings—but addressed tangentially or not at all by most previous frameworks—was that implementing health technology programs involves a great deal of work [106]. Normalization process theory unpacks implementation work into 4 categories: *coherence* (the work that people do to make sense of a practice), *cognitive participation* (work to enroll and engage other people in relation to that practice), *collective action* (work to enact the new practice, including efforts to bridge the model-practice gap described above), and *reflexive monitoring* (the work involved in assessing and adjusting a practice in use, including evaluating the impact of the technology and demonstrating its value to others) [27]. Implementation work may be particularly onerous in relation to health technologies because of the complexities and institutional challenges of addressing, for example, data security and patient privacy, interoperability across multiple information systems, resistance from health care professionals with a high degree of autonomy, and disruption to the critical granularity of clinical workflow [106]. First-order problems (such as slow technical performance) typically generate second-order problems, such as dramatically increased workloads (sometimes necessitating safety-critical workarounds), and perhaps third-order ones, such as reputational damage [106]. Failures in all 4 of May and Finch's work categories were evident in a systematic review of reasons for unsuccessful telemedicine implementation informed by normalization process theory [107].

Around half of the frameworks identified in this review included a question on the wider context for technology-supported care [108]. New technologies generate technical—and commercial and political—questions around interoperability standards, customer lock-in, customizability, substitutability, supplier relations, and the marketplace; policy questions about mainstreaming and funding new models of care; professional and managerial questions around standardization of algorithms and protocols; financial questions about who pays for which aspect of a networked service; legal questions around intellectual property; regulatory questions about safety, efficacy, and good business processes; and jurisdictional questions around liability, licensing, and the management of health information in shared environments [108-113]. The industry impetus of agile, rapid-iteration technology development and the "fail early, fail often" principle typically followed for software products contrasts with the risk-averse, highly regulated, and randomized trial-dominated context of much biomedical innovation [11,82,114].

While some implementation frameworks were designed around a rigid (and apparently systematic) logic model [51], we have previously argued that such an approach is counterproductive because eHealth technologies are typically introduced into a complex system in a turbulent and contested policy context [52]. Complexity principles distinguish simple phenomena (straightforward, predictable, few components) from complicated (multiple interacting components or issues) or complex (dynamic, unpredictable, not easily disaggregated into constituent components) ones [97]. Chambers et al's dynamic sustainability framework recognizes that, in order to be sustained, an innovation must adapt to its unique local environment and evolve over time [115], echoing Hawe et al's conceptualization of interventions as "events in [complex] systems" [116]. Several other recently published technology implementation frameworks have embraced complexity theory and argued strongly for a developmental, contextualized, and adaptive approach [43,47,53-60].

Abbott et al [60], for example, recommend that because the "same" program will play out differently in different contexts (and in the same context over time), collecting data at multiple levels, from multiple sites and longitudinally, will help elucidate these contextual influences. Because adaptation is key to embedding, inflexible milestones and overzealous measures of fidelity should be avoided. Local champions are likely to be key to participatory and developmental approaches; they should be identified early and partnered over time. Also key to the achievement of sustainability is attention to penetration (the extent to which the technology and its use become integrated with workflows so that workarounds are no longer necessary).

Van Gemert-Pijnen et al reviewed 16 frameworks for the implementation of eHealth innovations published up to 2009

(although in our own classification, 3 of these were not actually frameworks) [43]. They produced a multilevel theorization including diffusion of innovations, technology acceptance, service improvement, and organizational development. Their holistic technology implementation framework, known as the Centre for eHealth & Wellbeing Research roadmap, comprises 5 overlapping stages undertaken by a multidisciplinary team with iterative formative evaluation of each: (1) contextual inquiry (information gathering about the users and the environment, including ethnography and the use of scenarios); (2) value specification (including economic, social, and behavioral dimensions); (3) design (building prototypes that fit with values and user requirements); (4) operationalization (introduction and employment of the technology, including implementing a business model); and (5) summative evaluation (of uptake and impact).

Van Dyk, writing from an industrial engineering background, produced a taxonomy of eHealth implementation frameworks based on several theories, including diffusion of innovations, technology acceptance and use, e-readiness of organizations, transactional economics, and the information system life cycle, as well as considering more pragmatic barriers-and-facilitators studies [59]. She proposed a holistic approach that includes "technology, organizational structures, change management, economic feasibility, societal impacts, perceptions, user-friendliness, evaluation and evidence, legislation, policy and governance."

Van Dyk drew an important insight from organizational life cycle studies (including her own work on a telemedicine service maturity model for health care organizations [58])—that the challenges and constraints of technology-supported services vary with stage of development [59]. That is, (1) at the prototype stage, the main emphasis is on proof of concept and usability; (2) in small-scale pilots, it is on staff and societal acceptance (typically with an emphasis on the evidence base); (3) at the stage of wider local rollout—when financial support typically moves from external research grant to real set-up costs and a recurrent budget line—it is on financial and organizational considerations; and (4) when being considered for national rollout, it is on regulation, standardization, and security.

One final concept, which appears key to organizational adaptation over time but was not explicitly addressed in any of the previous frameworks, is the notion of resilience, defined as "the intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances so that it can sustain required operations, even after a major mishap or in the presence of continued stress" [117]. Organizational psychologists emphasize the importance of macrocognition: that is, reflecting collectively and continuously about how the organization is responding to change, including ongoing sense making, detecting critical events, and coordinating adaptive actions [118,119]. Introduction of new health care information technology systems can lead to loss of system resilience, since new technologies intended to automate work and assure safety may have the unintended effect of reducing time for collaborative dialogue, masking key trends in data (perhaps through information overload and loss of overview), making work routines brittle, and bypassing clinical judgment [118,120]. Cho et al have warned that the study of resilience in relation to new technologies requires multilevel analysis and is fraught with paradoxes (eg, that developing resilience in one part of the system may generate brittleness in another) [121].

In sum, we found the integrated frameworks of van Gemert-Pijnen et al [43] and Van Dyk [59] extremely helpful and used them as the starting point for analyzing our own dataset, modifying and refining them in the light of other high-quality frameworks published subsequently [53-55,60], and adding additional theoretical concepts (eg, burden of disease, health literacy, organizational resilience) and insights from our empirical data (especially our findings on the diverse and idiosyncratic nature of many conditions, which had received little attention in any previous framework).

Synthesis: The NASSS Framework

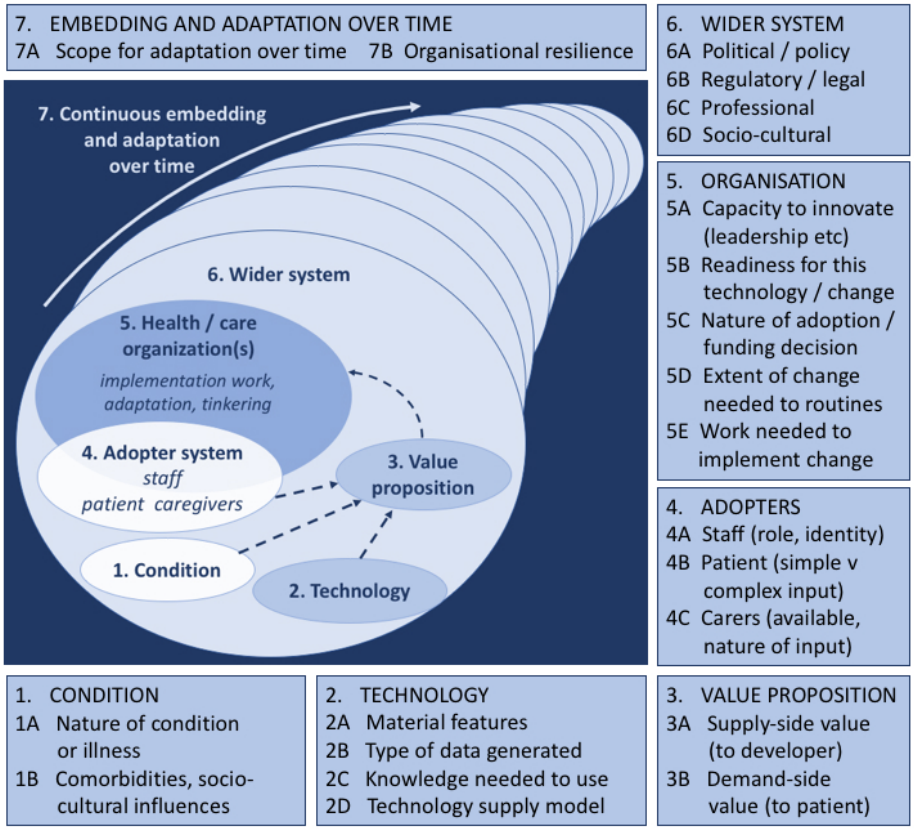
The final version of the NASSS technology implementation framework is shown in Figure 2 and expanded in Table 2. It consists of 13 questions in 6 domains: the condition, the technology, the value proposition, the adopter system (staff, patient, and lay caregiver[s]), the health or care organization(s) (including attention to the work of implementation and adaptation), and the wider (institutional and societal) context. It also includes a seventh domain that considers interactions and adaptations over time. The framework is intended to be used reflexively to guide conversations and help generate ideas, not as a checklist.

Our case studies raised a variety of challenges across all 7 domains, each of which could be classified as simple (straightforward, predictable, few components), complicated (multiple interacting components or issues), or complex (dynamic, unpredictable, not easily disaggregated into constituent components) [97]. Programs characterized by complicatedness proved difficult but not impossible to implement. Those characterized by complexity in multiple NASSS domains rarely, if ever, became mainstreamed. Multimedia Appendix 2 gives examples from our case studies in each of the domains, which we consider in turn.

The Condition

This domain addresses the clinical (question 1A) and the comorbidities and sociocultural aspects (question 1B) of the condition. It reflects the striking finding across all our case studies that only a fraction of potential end users were assessed by their clinicians as "suitable" for the technology. In the majority, the condition was considered clinically high risk, unpredictable, or atypical (eg, complicated by comorbidities or sociocultural factors, especially cognitive or health literacy considerations).

Figure 2. The NASSS framework for considering influences on the adoption, nonadoption, abandonment, spread, scale-up, and sustainability of patient-facing health and care technologies.



The Technology or Technologies

Question 2A addresses material and technical features of the technology. The technologies in our case studies were more or less freestanding. In all cases, features such as size, sounds, aesthetics, and "clunkiness" had significant impact on the technology's actual and perceived usability and appropriateness. Many were insufficiently prototyped. Dependability of technologies was key, especially in high-risk conditions or social situations.

Question 2B considers the knowledge generated or made visible by technology. This includes not only the accuracy of the data but also the extent to which those data are accepted, trusted, and considered sufficient for decision making. Engaging with the data generated by patient-facing technologies may inform, educate, and empower patients and lay caregivers. But data may also be misinterpreted by the patient or cause distress.

Question 2C addresses the knowledge and support needed to use the technology. Some technologies are much easier to operate than others; some require frequent troubleshooting; and some assume a different organizational role—or even an altered professional identity—for the user. Some patient-facing technologies require no knowledge form the patient; others require clinical knowledge, technical knowledge, and the ability to make judgments about (for example) what counts as urgent.

Question 2D addresses issues for sustainability raised by the technology supply model—that is, how the technology was procured, the nature of the client-supplier relationship, and the level of potential substitutability via the marketplace. The telehealth device marketplace is notorious for its lack of interoperability as companies attempt to retain their market share through lock-in of customers. The consequences for health and social care services—and their clients—of market withdrawal could be significant. While procurement options may not have immediate impact on scaling up and sustainability, they are likely to have significant impact in the long term and may influence decisions on whether to adopt an innovation in the first place. Customizable, off-the-shelf technologies (COTS) [113] offer relatively low cost (as they benefit from economies of scale) but may not be customizable to the extent that users really need (see domain 1). In contrast, bespoke solutions offer better fit with users' needs but at higher cost (both developmental and ongoing). Both options may be subject to risks of supplier withdrawal from the marketplace, with small and medium-sized enterprises being especially vulnerable.

The Value Proposition

This domain concerns whether a new technology is worth developing in the first place—and for whom it generates value. Question 3A addresses upstream value, which follows the supply-side logic of financial markets and investment decisions (and hence depends on preliminary tests of efficacy and safety, and evidence of good business practice). Question 3B addresses downstream value, which follows the demand-side logic of health technology appraisal, reimbursement, and procurement (ie, relates to evidence of benefit to patients and real-world affordability). As Multimedia Appendix 2 illustrates, our dataset included telling examples of mismatch between supply-side and demand-side value.

Table 2. Domains and questions in the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework.

Domain/question	Simple	Complicated	Complex
Domain 1: The condition or illness			
1A. What is the nature of the condition or illness?	Well-characterized, well-understood, predictable	Not fully characterized, understood, or predictable	Poorly characterized, poorly understood, unpredictable, or high risk
1B. What are the relevant sociocultural factors and comorbidities?	Unlikely to affect care significantly	Must be factored into care plan and service model	Pose significant challenges to care planning and service provision
Domain 2: The technology			
2A. What are the key features of the technology?	Off-the-shelf or already installed, freestanding, dependable	Not yet developed or fully interoperable; not 100% dependable	Requires close embedding in complex technical systems; significant dependability issues
2B. What kind of knowledge does the technology bring into play?	Directly and transparently measures [changes in] the condition	Partially and indirectly measures [changes in] the condition	Link between data generated and [changes in] the condition is currently unpredictable or contested
2C. What knowledge and/or support is required to use the technology?	None or a simple set of instructions	Detailed instruction and training needed, perhaps with ongoing helpdesk support	Effective use of technology requires advanced training and/or support to adjust to new identity or organizational role
2D. What is the technology supply model?	Generic, "plug and play," or COTS[a] solutions requiring minimal customization; easily substitutable if supplier withdraws	COTS solutions requiring significant customization or bespoke solutions; substitution difficult if supplier withdraws	Solutions requiring significant organizational reconfiguration or medium- to large scale-bespoke solutions; highly vulnerable to supplier withdrawal
Domain 3: The value proposition			
3A. What is the developer's business case for the technology (supply-side value)?	Clear business case with strong chance of return on investment	Business case underdeveloped; potential risk to investors	Business case implausible; significant risk to investors
3B. What is its desirability, efficacy, safety, and cost effectiveness (demand-side value)?	Technology is desirable for patients, effective, safe, and cost effective	Technology's desirability, efficacy, safety, or cost effectiveness is unknown or contested	Significant possibility that technology is undesirable, unsafe, ineffective, or unaffordable
Domain 4: The adopter system			
4A. What changes in staff roles, practices, and identities are implied?	None	Existing staff must learn new skills and/or new staff be appointed	Threat to professional identity, values, or scope of practice; risk of job loss
4B. What is expected of the patient (and/or immediate caregiver)—and is this achievable by, and acceptable to, them?	Nothing	Routine tasks, eg, log on, enter data, converse	Complex tasks, eg, initiate changes in therapy, make judgments, organize
4C. What is assumed about the extended network of lay caregivers?	None	Assumes a caregiver will be available when needed	Assumes a network of caregivers with ability to coordinate their input
Domain 5: The organization			
5A. What is the organization's capacity to innovate?	Well-led organization with slack resources and good managerial relations; risk taking encouraged	Limited slack resources; suboptimal leadership and managerial relations; risk taking not encouraged	Severe resource pressures (eg, frozen posts); weak leadership and managerial relations; risk taking may be punished
5B. How ready is the organization for this technology-supported change?	High tension for change, good innovation-system fit, widespread support	Little tension for change; moderate innovation-system fit; some powerful opponents	No tension for change; poor innovation-system fit; many opponents, some with wrecking power
5C. How easy will the adoption and funding decision be?	Single organization with sufficient resources; anticipated cost savings; no new infrastructure or recurrent costs required	Multiple organizations with partnership relationship; cost-benefit balance favorable or neutral; new infrastructure (eg, staff roles, training, kit) can mostly be found from repurposing	Multiple organizations with no formal links and/or conflicting agendas; funding depends on cost savings across system; costs and benefits unclear; new infrastructure conflicts with existing; significant budget implications

Domain/question	Simple	Complicated	Complex
5D. What changes will be needed in team interactions and routines?	No new team routines or care pathways needed	New team routines or care pathways that align readily with established ones	New team routines or care pathways that conflict with established ones
5E. What work is involved in implementation and who will do it?	Established shared vision; few simple tasks, uncontested and easily monitored	Some work needed to build shared vision, engage staff, enact new practices, and monitor impact	Significant work needed to build shared vision, engage staff, enact new practices, and monitor impact
Domain 6: The wider context			
6A. What is the political, economic, regulatory, professional (eg, medicolegal), and sociocultural context for program rollout?	Financial and regulatory requirements already in place nationally; professional bodies and civil society supportive	Financial and regulatory requirements being negotiated nationally; professional and lay stakeholders not yet committed	Financial and regulatory requirements raise tricky legal or other challenges; professional bodies and lay stakeholders unsupportive or opposed
Domain 7: Embedding and adaptation over time			
7A. How much scope is there for adapting and coevolving the technology and the service over time?	Strong scope for adapting and embedding the technology as local need or context changes	Potential for adapting and coevolving the technology and service is limited or uncertain	Significant barriers to further adaptation and/or coevolution of the technology or service
7B. How resilient is the organization to handling critical events and adapting to unforeseen eventualities?	Sense making, collective reflection, and adaptive action are ongoing and encouraged	Sense making, collective reflection, and adaptive action are difficult and viewed as low priority	Sense making, collective reflection, and adaptive action are discouraged in a rigid, inflexible implementation model

[a]COTS: customizable, off-the-shelf.

The Adopter System (Staff, Patient, Caregivers)

Question 4A is about adoption (and continued use) of the technology by staff. Reflecting the findings of previous studies reviewed above, some staff in each of our case studies simply did not engage with the program or use the technology. Nonadoption (or in some cases, abandonment) of the technology was occasionally explained by the technology's attributes (such as usability or ease of use). More commonly, staff were concerned about threats to their scope of practice or to the safety and welfare of the patient—and even, in some cases, about a fear of job loss.

Question 4B addresses adoption by patients or clients, including acceptance (hence symbolic meaning and aesthetics) and the work required of them.

Question 4C addresses the assumptions that may be built into the technology (or the linked service model) about the availability and behavior of lay caregivers. We encountered many cases of nonuse of all patient-facing technologies that were explained by weak or absent social networks, limited information technology skills (and distrust of technology) among lay caregivers, or long-standing family conflicts, which the technology sometimes brought to the surface but never solved (see examples in Multimedia Appendix 2).

The Organization(s)

Questions 5A and 5B address the organization's capacity (to embrace any service-level innovation) and readiness (for a specific technology), respectively. Our case studies included a wide range of antecedent conditions and levels of readiness that influenced uptake and internal scale-up of technology-supported programs (see examples in Multimedia Appendix 2).

Question 5C addresses the adoption decision, typically a board-level decision to allocate a budget line to support a particular technology. Our empirical data indicated three main problems with business modeling. The first problem was lack of data: it was often impossible to predict the uptake, use, and impact of the technology or the amount of investment needed to get (and keep) it up and running. Predictions were often guesstimates and typically overoptimistic, for example, in terms of numbers of potential users and potential costs and efficiency savings. The second problem was lack of money to support the program (see "organizational slack" above). Indeed, several organizations in our case studies did not appear to even recognize the need for a dedicated budget, over and above the cost of the technology, to support implementation and maintenance. The third problem was interdependencies between organizations and the teams working within and across them.

Question 5D considers the extent to which established work routines will be disrupted or made too brittle by the new technology. In some cases, there will be a transition period in which new collaborative routines linked to the new technology will be played out in parallel with existing routines—perhaps on legacy systems—before (hopefully) replacing them.

The last question in domain 5 concerns the *work* involved in implementation. All 6 case studies affirmed previous research that such work is extensive, often hidden and typically underestimated at the planning stage [122]. Our ethnographic data affirmed what Pols and Willems showed previously [29]: that technologies must be "tamed" in any particular setting by careful attention to the fine-grained detail of context (a process Pols and Willems called "tinkering" and Stewart and Williams called "domesticating" [105]). There is also the need for work on what Weick called sense making [95,103] and May and Finch called "coherence work" [27]: to make collective sense of the technology in the organizational setting and build a shared vision of its potential (including a realistic assessment of what the technology *cannot* do).

The Wider Context

Question 6A relates to the wider institutional and sociocultural context, which in our case studies was often key to explaining an organization's failure to move from a successful demonstration project (heavily dependent on particular champions and informal workarounds) to a fully mainstreamed service (scale-up) that was widely transferable (spread) and that persisted long term (sustainability). Aspects of wider context that proved pivotal in our case studies included health policy (including which service models were formally approved for funding—see example in Multimedia Appendix 2 of the difficulties in securing a nationally approved tariff for remote consultations), fiscal policy (the overall amount of funding available locally and nationally for health and care provision), the position taken by professional bodies and defense societies (who de facto defined what was acceptable professional practice), and legal and regulatory aspects of patient-facing technology development.

Interaction Between Domains and Adaptation Over Time

While the domains above can be distinguished analytically, the reality of any technology implementation project is that at an empirical level they are inextricably interlinked and dynamically evolving, often against a rapidly shifting policy context or continued evolution of the technology (and, at an individual patient level, as the condition deteriorates or fluctuates over time).

Question 7A relates to the medium- and long-term feasibility of continuing to adapt the technology and the program. Some technologies were more amenable to adaptation than others. Adaptation of staff roles and care pathways was also difficult when organizations were in subcontractor relationships (sometimes with private providers).

Question 7B concerns organizational resilience, and particularly what May and Finch called "reflexive monitoring" [27] and Patterson et al called the "macrocognitive functions" of sense making, including in particular the ability to detect critical events or issues and respond to these through coordinated action [118].

While an adaptive and reflexive approach is essential for effective scale-up, spread, and sustainability, it is impossible to be prescriptive at the outset as to how to go about this. Both formative and summative evaluation of the program's success will need to use imagination and multiple methods, along with a narrative account of what happened and why, in order to capture its multiple interdependencies, nonlinear effects, and unintended consequences [123].

Results of Prospectively Testing the NASSS Framework

Colleagues found our near-final NASSS framework helpful in considering challenges to implementation of a range of technology-supported health or care programs. While some used the framework (as we ourselves did) prospectively and in real time to predict and explore the challenges of implementing an existing technology to support a new program, some used it retrospectively to explain failures and partial successes in past projects. One group (a design company) suggested that the NASSS framework might be used at an early (design) stage to focus attention first on the condition and adopter system domains (ie, the needs of intended users), which would then inform both the technology and value proposition domains.

Discussion

Principal Results

Building on previous work by other groups, this study has developed and applied a new framework for predicting and evaluating the success of technology-supported health and social care programs. Across a diverse sample of 6 empirical case studies followed (so far) for up to 3 years and tested briefly (for what might be called face validity) in a further sample of 10 additional case studies, we identified and explained numerous examples of nonadoption and abandonment of the technology by individuals or limited success in attempts to scale up, spread, and sustain the program within and beyond the organization.

A striking finding that explained many such instances was a tendency to assume that the issues to be addressed were simple or complicated (hence knowable, predictable, and controllable) rather than *complex* (that is, inherently not knowable or predictable but dynamic and emergent). Common problems included the following:

- The technology and linked program had been designed around an oversimplified, textbook model of the condition.
- The technology was insufficiently prototyped, insufficiently customizable, insufficiently dependable, dependent on complex knowledge to use it, or designed to generate data or knowledge that was incomplete, inappropriate, or contested in the care context.
- The value proposition of the technology was unclear, in terms of a viable business venture for its developer or in terms of a clear benefit for patients and an affordable real-world service model.
- The intended users of the technology had plausible personal or professional reasons to resist or reject it.
- The organization(s) were inadequately set up for innovation, not ready for (or interested in) this particular innovation, unable to negotiate a viable business model with partner organizations, unable to shift to new ways of working, or unable to support the extensive work needed to implement and sustain the change (including the rhetorical "work" of making sense of an equivocal technology).
- Complexity in external (financial, regulatory, legal, policy) issues—of which reimbursement seemed to be particularly key—stalled the mainstreaming and spread of the program.
- The program was unable to adapt and evolve over time in a way that continued to meet the needs of its intended users and remain clinically, operationally, and financially viable.

Strengths and Limitations

The NASSS framework has been developed systematically to fill a key gap in the literature on technology implementation—specifically, to address not just adoption but also *nonadoption* and *abandonment* of technologies and the challenges associated with moving from a local demonstration

project to one that is fully mainstreamed and part of business as usual locally (scale-up), transferable to new settings (spread), and maintained long term through adaptation to context over time (sustainability).

NASSS was never intended to provide a simple fix for complex problems. It cannot be applied in a formulaic way or used deterministically as a tool, nor do we believe it is possible to make firm predictions about which elements of the framework will be mission-critical or how the different elements will interact (since these are likely to differ substantially in different cases and settings). With these caveats, we believe that, subject to further empirical testing, the NASSS diagram in Figure 2 and the NASSS domains and questions in Table 2 have several potential uses. In particular, they could be used (1) at an early stage in a technology development to inform technology and service design; (2) in strategic planning to identify technological innovations that (perhaps despite policy enthusiasm) have limited chance of achieving large-scale, sustained adoption; (3) in technology implementation projects to address the micro-level challenges of individual adoption, the meso-level challenges of organizational assimilation, and the macro-level challenges of the policy and regulatory environment; (4) to inform and support scale-up and rollout of technology programs; and (5) retrospectively to explain program failures.

The empirical work to develop NASSS was based entirely in the United Kingdom, although collaborators who peer reviewed and tested the draft framework included groups from Australia, Canada, Italy, and the United States. The 6 case studies presented in this paper have been followed for up to 3 years; field work is ongoing, so additional insights about long-term sustainability may emerge in the future. We strongly encourage other research groups to explore the applicability of the framework for different purposes and to adapt and extend it if appropriate.

Comparison With Previous Work

Previous technology implementation frameworks identified in our hermeneutic review of the literature are summarized in Multimedia Appendix 1; the approaches and limitations of those frameworks are described in the Results section. We selected 2 rigorously developed, multilevel, integrative frameworks as most closely fitting our own empirical data [43,59] and have built on these in three main ways.

First, we have added a preliminary focus on the illness or condition for which the technology is assumed to be a solution, an emphasis that emerged from our empirical data but probably also reflects the background of several authors as clinicians, psychologists, or patient-facing service coordinators. Second, we have introduced the heuristic of classifying each domain of interest as simple, complicated, or complex, and cautiously concluded that it is *complexity in multiple domains* that poses the greatest challenge to scale-up, spread, and sustainability. Third, while acknowledging our academic audiences and not wishing to oversimplify a problem that is inherently complex, we have produced a visual representation of the different NASSS domains that we hope will be accessible to key nonacademic audiences: clinicians, managers, technology developers, executive decision makers in health and care organizations, and patients and caregivers.

Conclusions

Implementing new technologies as part of changes to health and social care services is inherently challenging. While policy makers are calling for technology to be implemented rapidly and at scale, the reality is that when dealing with the multiple complexities of health and care, it is extremely difficult to go beyond small-scale demonstration projects. We hope that the NASSS framework will help implementation teams—and, at an earlier stage, technology and service designers—to identify, understand, and address the interacting challenges to achieving sustained adoption, local scale-up, distant spread, and long-term sustainability of their programs.

While we believe that the NASSS framework is academically defensible, additional work is now ongoing to make it accessible to its intended users beyond academia. In the United Kingdom, the NHS National Technology Adoption Centre (NTAC) was set up in 2007 to promote the uptake of health technologies through its Technology Implementation Projects and HowToWhyTo guides (which attempt to distil lessons learned from pilots in a format designed to be easily transferable and hence reusable). One of the principal objectives of the guides has been to assist practitioner stakeholders to build a business case for innovation. However, a recent study suggests that their impact has been disappointing and points to the difficulties of making context-specific knowledge transferable and hence reusable in other contexts [124]. Cognizant of this, we are currently working with design colleagues to develop and evaluate accessible infographic summaries, Web tools, and team learning opportunities based on the NASSS framework for different audiences.

Acknowledgments

We are grateful to the participants (patients, caregivers, and staff) in the VOCAL and SCALS studies for their commitment and insights. We also thank the study steering groups, especially Mrs Jenni Bowley who chaired them. We are grateful to various colleagues who tested the NASSS framework on their own projects and provided feedback, and especially staff at Frog Design who contributed to substantial changes to the NASSS diagram. We drew on a qualitative evaluation of Product B undertaken by researchers at another university and published internally by the caregivers' charity.

The VOCAL study was funded through the UK National Institute for Health Research Health Services and Delivery Research grant to TG (13/59/26). The SCALS program's main funding is from a Senior Investigator Award and Public Engagement Award to TG from the Wellcome Trust in its Society and Ethics Program (WT104830MA). Additional funding for the GPS tracking

case study was provided through a Program Development Grant to TG from the National Institute for Health Research (RP-DG-1213-10003). The SCALS study is registered on the Wellcome Trust's Society and Ethics portfolio. TG, SS, JW, and NF are part-funded by the National Institute for Health Research Biomedical Research Centre, Oxford, UK (grant NIHR-BRC-1215-20008). The initial draft of this paper was written when TG was on a writing retreat at the Bellagio Center, Italy, funded by the Rockefeller Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Previous implementation/evaluation frameworks for health and care technologies.

[PDF File (Adobe PDF File), 47KB - jmir_v19i11e367_app1.pdf]

Multimedia Appendix 2

Examples from case studies to illustrate NASSS.

[PDF File (Adobe PDF File), 63KB - jmir_v19i11e367_app2.pdf]

References

1. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
2. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. Soc Sci Med 2005 Jul;61(2):417-430. [doi: 10.1016/j.socscimed.2004.12.001] [Medline: 15893056]
3. Robert G, Greenhalgh T, MacFarlane F, Peacock R. Adopting and assimilating new non-pharmaceutical technologies into health care: a systematic review. J Health Serv Res Policy 2010 Oct;15(4):243-250. [doi: 10.1258/jhsrp.2010.009137] [Medline: 20592046]
4. Greenhalgh T. Role of routines in collaborative work in healthcare organisations. BMJ 2008 Nov 17;337:a2448. [Medline: 19015186]
5. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
6. Chaudoir SR, Dugan AG, Barr CH. Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. Implement Sci 2013 Feb 17;8:22 [FREE Full text] [doi: 10.1186/1748-5908-8-22] [Medline: 23414420]
7. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
8. De Vries H, Bekkers V, Tummers L. Innovation in the public sector: a systematic review and future research agenda. Public Adm 2016;94(1):146-166.
9. Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. National Institutes of Health approaches to dissemination and implementation science: current and future directions. Am J Public Health 2012 Jul;102(7):1274-1281. [doi: 10.2105/AJPH.2012.300755] [Medline: 22594758]
10. Garber S, Gates S, Keeler E, Valana M, Mulcahy A, Lau C. Redirecting innovation in U.S. health care: options to decrease spending and increase value. Santa Monica, CA: RAND Corporation; 2014. URL: https://www.rand.org/pubs/research_reports/RR308.html [accessed 2017-10-19] [WebCite Cache ID 6uL1hXoj5]
11. van Limburg M, van Gemert-Pijnen JEWC, Nijland N, Ossebaard HC, Hendrix RMG, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. J Med Internet Res 2011 Dec 28;13(4):e124 [FREE Full text] [doi: 10.2196/jmir.1674] [Medline: 22204896]
12. Grin J, Rotmans J, Schot J. Transitions to Sustainable Development: New Directions in the Study of Long Term Transformative Change. New York, NY: Routledge; 2010.
13. Greenhalgh T, A'Court C, Shaw S. Understanding heart failure; explaining telehealth - a hermeneutic systematic review. BMC Cardiovasc Disord 2017 Jun 14;17(1):156 [FREE Full text] [doi: 10.1186/s12872-017-0594-2] [Medline: 28615004]
14. Standing C, Standing S, McDermott M, Gururajan R, Kiani Mavi R. The paradoxes of telehealth: a review of the literature 2000-2015. Syst Res 2016 Dec 01. [doi: 10.1002/sres.2442]

15. Bentley C, Powell L, Orrell A, Mountain G. Addressing design and suitability barriers to telecare use: has anything changed? Technol Disabil 2014;26(4):221-235. [doi: 10.3233/TAD-150421]
16. Clark J, McGee-Lennon M. A stakeholder-centred exploration of the current barriers to the uptake of home care technology in the UK. J Assist Technol 2011;5(1):12-25. [doi: 10.5042/jat.2011.0097]
17. Wade V, Eliott J, Hiller J. Clinician acceptance is the key factor for sustainable telehealth services. Qual Health Res 2014;24(5):682-694. [doi: 10.4973/2314528809]
18. Zanaboni P, Wootton R. Adoption of routine telemedicine in Norwegian hospitals: progress over 5 years. BMC Health Serv Res 2016;16(1):496. [Medline: 27644324]
19. Sligo J, Gauld R, Roberts V, Villa L. A literature review for large-scale health information system project planning, implementation and evaluation. Int J Med Inform 2017;97:86-97. [Medline: 27919399]
20. Gagnon M, Nsangou ER, Payne-Gagnon J, Grenier S, Sicotte C. Barriers and facilitators to implementing electronic prescription: a systematic review of user groups? perceptions. J Am Med Inform Assoc 2014;21(3):a. [Medline: 24130232]
21. Hogan-Murphy D, Tonna A, Strath A, Cunningham S. Healthcare professionals' perceptions of the facilitators and barriers to implementing electronic systems for the prescribing, dispensing and administration of medicines in hospitals: a systematic review. Eur J Hosp Pharm 2015;22(6):358-365. [doi: 10.1136/ejhpharm-2015-000722]
22. Bergs J, Lambrechts F, Simons P, Vlayen A, Marneffe W, Hellings J, et al. Barriers and facilitators related to the implementation of surgical safety checklists: a systematic review of the qualitative evidence. BMJ Qual Saf 2015 Dec;24(12):776-786. [doi: 10.1136/bmjqs-2015-004021] [Medline: 26199428]
23. Atun R, de Jongh T, Secci F, Ohiri K, Adeyi O. Integration of targeted health interventions into health systems: a conceptual framework for analysis. Health Policy Plan 2010 Mar;25(2):104-111. [doi: 10.1093/heapol/czp055] [Medline: 19917651]
24. Kessler R, Glasgow RE. A proposal to speed translation of healthcare research into practice: dramatic change is needed. Am J Prev Med 2011 Jun;40(6):637-644. [doi: 10.1016/j.amepre.2011.02.023] [Medline: 21565657]
25. Greenhalgh T, Shaw S, Wherton J, Hughes G, Lynch J, A'Court C, et al. SCALS: a fourth-generation study of assisted living technologies in their organisational, social, political and policy context. BMJ Open 2016 Feb 15;6(2):e010208 [FREE Full text] [doi: 10.1136/bmjopen-2015-010208] [Medline: 26880671]
26. Lupton D. Critical perspectives on digital health technologies. Sociol Compass 2014 Dec 04;8(12):1344-1359. [doi: 10.1111/soc4.12226]
27. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
28. Nicolini D. Practice as the site of knowing: insights from the field of telemedicine. Organ Sci 2011 Jun;22(3):602-620. [doi: 10.1287/orsc.1100.0556]
29. Pols J, Willems D. Innovation and evaluation: taming and unleashing telecare technology. Sociol Health Illn 2011 Mar;33(3):484-498 [FREE Full text] [doi: 10.1111/j.1467-9566.2010.01293.x] [Medline: 21241338]
30. Maniatopoulos G, Procter R, Llewellyn S, Harvey G, Boyd A. Moving beyond local practice: reconfiguring the adoption of a breast cancer diagnostic technology. Soc Sci Med 2015 Apr;131:98-106. [doi: 10.1016/j.socscimed.2015.02.036] [Medline: 25769108]
31. Greenhalgh T, Stones R. Theorising big IT programmes in healthcare: strong structuration theory meets actor-network theory. Soc Sci Med 2010 May;70(9):1285-1294. [doi: 10.1016/j.socscimed.2009.12.034] [Medline: 20185218]
32. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016 Jan 29;6(1):e009388. [doi: 10.1136/bmjopen-2015-009388]
33. Gorst SL, Armitage CJ, Brownsell S, Hawley MS. Home telehealth uptake and continued use among heart failure and chronic obstructive pulmonary disease patients: a systematic review. Ann Behav Med 2014 Dec;48(3):323-336 [FREE Full text] [doi: 10.1007/s12160-014-9607-x] [Medline: 24763972]
34. Brewster L, Mountain G, Wessels B, Kelly C, Hawley M. Factors affecting front line staff acceptance of telehealth technologies: a mixed-method systematic review. J Adv Nurs 2014 Jan;70(1):21-33. [doi: 10.1111/jan.12196] [Medline: 23786584]
35. Boyne JJJ, Vrijhoef HJM. Implementing telemonitoring in heart failure care: barriers from the perspectives of patients, healthcare professionals and healthcare organizations. Curr Heart Fail Rep 2013 Sep;10(3):254-261. [doi: 10.1007/s11897-013-0140-1] [Medline: 23666901]
36. Merrell RC, Doarn CR. Barriers or barricades. Telemed J E Health 2012 Mar;18(2):79-80. [doi: 10.1089/tmj.2012.9997] [Medline: 22381059]
37. Hunting G, Shahid N, Sahakyan Y, Fan I, Moneypenny CR, Stanimirovic A, et al. A multi-level qualitative analysis of Telehomecare in Ontario: challenges and opportunities. BMC Health Serv Res 2015 Dec 09;15:544 [FREE Full text] [doi: 10.1186/s12913-015-1196-2] [Medline: 26645639]
38. Lluch M, Abadie F. Exploring the role of ICT in the provision of integrated care--evidence from eight countries. Health Policy 2013 Jun;111(1):1-13. [doi: 10.1016/j.healthpol.2013.03.005] [Medline: 23587547]
39. Gheorghiu B, Ratchford F. Scaling up the use of remote patient monitoring in Canada. Stud Health Technol Inform 2015;209:23-26. [Medline: 25980701]

40. Saner H, van der Velde E. eHealth in cardiovascular medicine: a clinical update. Eur J Prev Cardiol 2016 Oct;23(2 suppl):5-12. [doi: 10.1177/2047487316670256] [Medline: 27892420]
41. Dierckx R, Pellicori P, Cleland JGF, Clark AL. Telemonitoring in heart failure: Big Brother watching over you. Heart Fail Rev 2015 Jan;20(1):107-116. [doi: 10.1007/s10741-014-9449-4] [Medline: 24972644]
42. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ 2005 Nov 05;331(7524):1064-1065 [FREE Full text] [doi: 10.1136/bmj.38636.593461.68] [Medline: 16230312]
43. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
44. Lynch J. Policy Aspirations and Practice in English Telecare: A Case Study of Story-lines and Invisible Work (PhD thesis). Birmingham, UK: University of Birmingham; 2015.
45. Procter R, Wherton J, Greenhalgh T, Sugarhood P, Rouncefield M, Hinder S. Telecare call centre work and ageing in place. Comput Support Coop Work 2016 Jan 15;25(1):79-105. [doi: 10.1007/s10606-015-9242-5]
46. Triantafyllidis A, Velardo C, Chantler T, Shah SA, Paton C, Khorshidi R, et al. A personalised mobile-based home monitoring system for heart failure: the SUPPORT-HF Study. Int J Med Inform 2015 Oct;84(10):743-753. [doi: 10.1016/j.ijmedinf.2015.05.003] [Medline: 26037921]
47. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Hinder S, Rouncefield M. What is quality in assisted living technology? The ARCHIE framework for effective telehealth and telecare services. BMC Med 2015;13:91 [FREE Full text] [doi: 10.1186/s12916-015-0279-6] [Medline: 25902803]
48. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
49. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for health information systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008 Jun;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
50. Van Der Meijden MJ, Tange HJ, Troost J, Hasman A. Determinants of success of inpatient clinical information systems: a literature review. J Am Med Inform Assoc 2003;10(3):235-243 [FREE Full text] [doi: 10.1197/jamia.M1094] [Medline: 12626373]
51. Chang H. Evaluation framework for telemedicine using the logical framework approach and a fishbone diagram. Healthc Inform Res 2015 Oct;21(4):230-238 [FREE Full text] [doi: 10.4258/hir.2015.21.4.230] [Medline: 26618028]
52. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
53. Anton MT, Jones DJ. Adoption of technology - enhanced treatments: conceptual and practical considerations. Clin Psychol (New York) 2017 Sep;24(3):223-240. [doi: 10.1111/cpsp.12197] [Medline: 28966479]
54. Madjedi K, Daya R. Towards a framework for the development, implementation and sustainability of eHealth interventions in Indigenous communities. Univ Ottawa J Med 2016;6(1):27-30.
55. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
56. Agboola S, Hale TM, Masters C, Kvedar J, Jethwani K. "Real-world" practical evaluation strategies: a review of telehealth evaluation. JMIR Res Protoc 2014;3(4):e75 [FREE Full text] [doi: 10.2196/resprot.3459] [Medline: 25524892]
57. van Dyk L, Schutte C. The telemedicine service maturity model: a framework for the measurementimprovement of telemedicine services. In: Madhaavn R, Khalid S, editors. Telemedicine. London, UK: Intech; 2013.
58. Kukafka R, Johnson SB, Linfante A, Allegrante JP. Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform 2003 Jun;36(3):218-227 [FREE Full text] [Medline: 14615230]
59. van Dyk L. A review of telehealth service implementation frameworks. Int J Environ Res Public Health 2014 Jan 23;11(2):1279-1298 [FREE Full text] [doi: 10.3390/ijerph110201279] [Medline: 24464237]
60. Abbott PA, Foster J, Marin HDF, Dykes PC. Complexity and the science of implementation in health IT--knowledge gaps and future visions. Int J Med Inform 2014 Jul;83(7):e12-e22. [doi: 10.1016/j.ijmedinf.2013.10.009] [Medline: 24444700]
61. Eng TR, Gustafson DH, Henderson J, Jimison H, Patrick K. Introduction to evaluation of interactive health communication applications. Science Panel on Interactive Communication and Health. Am J Prev Med 1999 Jan;16(1):10-15. [Medline: 9894549]
62. Hebert M. Telehealth success: evaluation framework development. Stud Health Technol Inform 2001;84(Pt 2):1145-1149. [Medline: 11604908]
63. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]

64. Shaw NT. 'CHEATS': a generic information communication technology (ICT) evaluation framework. Comput Biol Med 2002 May;32(3):209-220. [Medline: 11922936]
65. Kazanjian A, Green CJ. Beyond effectiveness: the evaluation of information systems using a comprehensive health technology assessment framework. Comput Biol Med 2002 May;32(3):165-177. [Medline: 11922933]
66. Ganesh J. E-health drivers, applications, challenges ahead and Strategies: a conceptual framework. Indian J Med Inform 2004;1(1):40-48.
67. Dansky KH, Thompson D, Sanner T. A framework for evaluating eHealth research. Eval Program Plann 2006 Nov;29(4):397-404. [doi: 10.1016/j.evalprogplan.2006.08.009] [Medline: 17950868]
68. Hamid A, Sarmad A. Evaluation of e‐health services: user's perspective criteria. Transform Gov 2008 Aug;2(4):243-255. [doi: 10.1108/17506160810917945]
69. Esser PE, Goossens RHM. A framework for the design of user-centred teleconsulting systems. J Telemed Telecare 2009;15(1):32-39. [doi: 10.1258/jtt.2008.080601] [Medline: 19139218]
70. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med 2002 May;32(3):141-149. [Medline: 11922931]
71. Kaufman D, Roberts WD, Merrill J, Lai T, Bakken S. Applying an evaluation framework for health information system design, development, and implementation. Nurs Res 2006;55(2 Suppl):S37-S42. [Medline: 16601633]
72. Wouters E, Weijers T, Finch T. Successful implementation of technological innovations in health care organizations. In: van Hoof J, Demiris G, Wouters EJM, editors. Handbook of Smart Homes, Health Care and Well-Being. Amsterdam, Netherlands: Springer; 2017:179-189.
73. Tait GR, Bates J, LaDonna KA, Schulz VN, Strachan PH, McDougall A, et al. Adaptive practices in heart failure care teams: implications for patient-centered care in the context of complexity. J Multidiscip Healthc 2015;8:365-376 [FREE Full text] [doi: 10.2147/JMDH.S85817] [Medline: 26316775]
74. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012;2(4):001574 [FREE Full text] [doi: 10.1136/bmjopen-2012-001574] [Medline: 22815469]
75. May CR, Eton DT, Boehmer K, Gallacher K, Hunt K, MacDonald S, et al. Rethinking the patient: using burden of treatment theory to understand the changing dynamics of illness. BMC Health Serv Res 2014;14:281 [FREE Full text] [doi: 10.1186/1472-6963-14-281] [Medline: 24969758]
76. Showell C, Cummings E, Turner P. The invisibility of disadvantage: why do we not notice? Stud Health Technol Inform 2017;235:388-392. [Medline: 28423820]
77. World Health Organization. Social Determinants of Health. Geneva, Switzerland: WHO; 2017 Jan 01. URL: http://www.who.int/social_determinants/en/ [accessed 2017-10-19] [WebCite Cache ID 6uKmPaYzD]
78. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
79. Pols J. Care at a Distance: On the Closeness of Technology. Amsterdam, Netherlands: Amsterdam University Press; 2012.
80. Rogers E. Diffusion of Innovations. New York, NY: Simon and Schuster; 2010.
81. Lehoux P. The duality of health technology in chronic illness: how designers envision our future. Chronic Illn 2008 Jun;4(2):85-97. [doi: 10.1177/1742395308092475] [Medline: 18583444]
82. Lehoux P, Miller FA, Daudelin G, Denis J. Providing value to new health technology: the early contribution of entrepreneurs, investors, and regulatory agencies. Int J Health Policy Manag 2017 Jan 25;6(9):509-518 [FREE Full text] [doi: 10.15171/ijhpm.2017.11] [Medline: 28949463]
83. Lehoux P, Miller FA, Daudelin G, Urbach DR. How venture capitalists decide which new medical technologies come to exist. Sci Public Policy 2015 Aug 27;43(3):375-385. [doi: 10.1093/scipol/scv051]
84. Greenhalgh T, Shaw S, Fahy N. The bright elusive butterfly of value in health technology development. Int J Health Policy Manage 2017;6:1-5. [doi: 10.15171/ijhpm.2017.65]
85. Markiewicz K, van Til JA, IJzerman MJ. Medical devices early assessment methods: systematic literature review. Int J Technol Assess Health Care 2014 Apr;30(2):137-146. [doi: 10.1017/S0266462314000026] [Medline: 24805836]
86. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
87. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice-Hall; 1976.
88. Williams R, Stewart J, Slack R. Social Learning in Technological Innovation: Experimenting With Information and Communication Technologies. Cheltenham, UK: Edward Elgar Publishing; 2005.
89. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. Inf Manage 2003 Jan;40(3):191-204. [doi: 10.1016/S0378-7206(01)00143-4]
90. Sharma U, Clarke M. Nurses' and community support workers' experience of telehealth: a longitudinal case study. BMC Health Serv Res 2014 Apr 10;14:164 [FREE Full text] [doi: 10.1186/1472-6963-14-164] [Medline: 24721599]
91. Taylor J, Coates E, Brewster L, Mountain G, Wessels B, Hawley MS. Examining the use of telehealth in community nursing: identifying the factors affecting frontline staff acceptance and telehealth adoption. J Adv Nurs 2015 Feb;71(2):326-337. [doi: 10.1111/jan.12480] [Medline: 25069605]

92. Wade V, Eliott J. The role of the champion in telehealth service development: a qualitative analysis. J Telemed Telecare 2012 Dec;18(8):490-492. [doi: 10.1258/jtt.2012.gth115] [Medline: 23209264]
93. Greenhalgh T, Swinglehurst D, Stones R. Rethinking resistance to 'big IT': a sociological study of why and when healthcare staff do not use nationally mandated information and communication technologies. Health Serv Deliv Res 2014 Nov;39(2):1-86. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 27466649]
94. DeLone WH, McLean ER. Information systems success: the quest for the dependent variable. Inf Syst Res 1992 Mar;3(1):60-95. [doi: 10.1287/isre.3.1.60]
95. Weick K. Technology as equivoque: sensemaking in new technologies. In: Goodman PS, Sproull LS, editors. Technology and Organizations. San Francisco, CA: Jossey-Bass; 1990:1-44.
96. Zahra SA, George G. Absorptive capacity: a review, reconceptualization, and extension. Acad Manage Rev 2002 Apr 01;27(2):185-203. [doi: 10.5465/AMR.2002.6587995]
97. Plsek PE, Greenhalgh T. Complexity science: the challenge of complexity in health care. BMJ 2001 Sep 15;323(7313):625-628 [FREE Full text] [Medline: 11557716]
98. Feldman MS. Organizational routines as a source of continuous change. Organ Sci 2000 Dec;11(6):611-629. [doi: 10.1287/orsc.11.6.611.12529]
99. Ackerman M. The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. Hum Comput Interact 2000 Sep 1;15(2):179-203. [doi: 10.1207/S15327051HCI1523_5]
100. Symon G, Long K, Ellis J. The coordination of work activities: cooperation and conflict in a hospital context. Comput Support Coop Work 1996 Mar;5(1):1-31. [doi: 10.1007/BF00141934]
101. Cherns A. Principles of sociotechnical design revisted. Hum Relat 2016 Apr 22;40(3):153-161. [doi: 10.1177/001872678704000303]
102. 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]
103. Weick K. Sensemaking in Organizations. Thousand Oaks, CA: Sage; 1995.
104. Orlikowski WJ. The duality of technology: rethinking the concept of technology in organizations. Organ Sci 1992 Aug;3(3):398-427. [doi: 10.1287/orsc.3.3.398]
105. Stewart J, Williams R. The wrong trousers? Beyond the design fallacy: social learning and the user. In: Rohracher H, editor. User Involvement in Innovation Processes. Strategies and Limitations From a Socio-Technical Perspective. Munich, Germany: Profil-Verlag; 2005.
106. Doolin B. Implementing e-Health. In: Ferlie E, Montgomery K, Reff Pedersen A, editors. The Oxford Handbook of Health Care Management. Oxford, UK: Oxford University Press; 2016:352-374.
107. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 1;90(5):357-364 [FREE Full text] [doi: 10.2471/BLT.11.099424] [Medline: 22589569]
108. Khoja S, Durrani H, Nayani P, Fahim A. Scope of policy issues in eHealth: results from a structured literature review. J Med Internet Res 2012 Feb 17;14(1):e34 [FREE Full text] [doi: 10.2196/jmir.1633] [Medline: 22343270]
109. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet 2011 Aug 20;378(9792):731-739. [doi: 10.1016/S0140-6736(11)61229-4] [Medline: 21856487]
110. Siegal G. Telemedicine: licensing and other legal issues. Otolaryngol Clin North Am 2011 Dec;44(6):1375-84, xi. [doi: 10.1016/j.otc.2011.08.011] [Medline: 22032489]
111. Wismar M, Palm W, Figueras J, Ernst K, van Ginneken E. Cross-Border Health Care in the European Union. Copenhagen, Denmark: WHO Europe; 2011.
112. Williams R. Universal solutions or local contingencies? Tensions and contradictions in the mutual shaping of technology work in organisations. In: McLoughlin I, Preace D, Dawson P, editors. Technology, Organizations and Innovation. Volume 3: Critical Empirical Studies. London, UK: Routledge; 2000:2000-2039.
113. Pollock N, Williams R, Procter R. Fitting standard software packages to non-standard organizations: the 'biography' of an enterprise-wide system. Technol Anal Strateg Manage 2003 Sep;15(3):317-332. [doi: 10.1080/0953732031000160150 4]
114. Van Norman GA. Drugs and devices. JACC Basic Transl Sci 2016 Aug;1(5):399-412. [doi: 10.1016/j.jacbts.2016.06.003]
115. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013 Oct 02;8:117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]
116. Hawe P, Shiell A, Riley T. Theorising interventions as events in systems. Am J Community Psychol 2009 Jun;43(3-4):267-276. [doi: 10.1007/s10464-009-9229-9] [Medline: 19390961]
117. Nemeth C, Wears R, Woods D, Hollnagel E, Cook R. Minding the gaps: creating resilience in health care. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in Patient Safety: New Directions and Alternative Approaches. Volume 3: Performance and Tools. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
118. Patterson ES, Militello LG, Su G, Sarkar U. Characterizing a naturalistic decision-making phenomenon: loss of system resilience associated with implementation of new technology. J Cogn Eng Decis Mak 2016 Sep;10(3):229-243 [FREE Full text] [doi: 10.1177/1555343416652524] [Medline: 28138316]

119. Procter R, Rouncefield M, Balka E, Berg M. Special issue: CSCW and dependable healthcare systems. Comput Support Coop Work 2007 Jan 6;15(5-6):413-418. [doi: 10.1007/s10606-006-9037-9]
120. Jirotka M, Procter R, Hartswood M, Slack R, Simpcson A, Coopmans C, et al. Collaboration and trust in healthcare innovation: the eDiaMoND case study. Comput Support Coop Work 2005 Sep 14;14(4):369-398. [doi: 10.1007/s10606-005-9001-0]
121. Cho S, Mathiassen L, Robey D. The dialectics of resilience: a multilevel analysis of a telehealth innovation. In: Donnellan B, Larsen TJ, Levine L, DeGross JI, editors. The TDIT 2006: Transfer and Diffusion of Information Technology for Organizational Resilience. Cham Switzerland: Springer International Publishing AG; 2006:339-357.
122. Oudshoorn N. Diagnosis at a distance: the invisible work of patients and healthcare professionals in cardiac telemonitoring technology. Sociol Health Illn 2008 Mar;30(2):272-288 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01032.x] [Medline: 18290936]
123. Greenhalgh T, Macfarlane F, Barton-Sweeney C, Woodard F. "If we build it, will it stay?" A case study of the sustainability of whole-system change in London. Milbank Q 2012 Sep;90(3):516-547 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00673.x] [Medline: 22985280]
124. Llewellyn S, Procter R, Harvey G, Maniatopoulos G, Boyd A. Facilitating technology adoption in the NHS: negotiating the organisational and policy context-a qualitative study. 2014. Health Serv Deliv Res 2014 Jul;2:23. [Medline: 25642495]

Abbreviations

ARCHIE: anchored, realistic, cocreative, human, integrated, and evaluated
COTS: customizable, off-the-shelf
GPS: global positioning system
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
SCALS: Studies in Co-creating Assisted Living Solutions
SUPPORT-HF: Seamless User-Centred Proactive Provision of Risk-Stratified Treatment for Heart Failure
VOCAL: Virtual Online Consultations: Advantages and Limitations

Edited by G Eysenbach; submitted 17.08.17; peer-reviewed by C May, B Gheorghiu, E Cummings; comments to author 07.09.17; revised version received 08.09.17; accepted 23.09.17; published 01.11.17.

Please cite as:
Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, Hinder S, Fahy N, Procter R, Shaw S
Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies
J Med Internet Res 2017;19(11):e367
URL: http://www.jmir.org/2017/11/e367/
doi:10.2196/jmir.8775
PMID:29092808

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org